Probability on Graphs
Summary of Lectures

MCS 494 Special Topics in Computer Science, Spring 2005,

20382 LCD - undergrad, 20384 - grad, MWF 3:00-3:50, SES 170

Instructor: Shmuel Friedland
Department of Mathematics, Statistics and Computer Science,
email: friedlan@uic.edu

Last update April 4, 2005

1 Probability

1.1 Events

Let Ω be a *sample space*. We are going to assume that Ω is either a finite set $\{\omega_1, \dots, \omega_n\}$, or infinite countable set $\{\omega_i, i \in \mathbb{N}\}$, where $\mathbb{N} = \{1, \dots, \}$ is the set of positive integers. The two kinds of such sets are called *countable*. Denote by $\#\Omega$ the *cardinality* of Ω. Thus $\#\Omega = n \in \mathbb{N}$ if Ω is a finite set. $\#\Omega = \aleph_0$ if Ω is infinite countable. To each $\omega \in \Omega$ we attach a probability (mass) $p(\omega) \geq 0$. The *normalization condition* is $\sum_{\omega \in \Omega} p(\omega) = 1$. That is, the total mass of Ω is 1. The row vector $\mu := (p(\omega_1), p(\omega_2), \dots)$ is also called sometimes a *probability measure* on Ω, or *distribution*.

A subset A of Ω, denoted as $A \subset \Omega$ is called an *event*. The set of all subsets of Ω is denoted by 2^Ω. It includes the *empty set*, denoted by $\emptyset$ and Ω. Then $\Pr(A) := \sum_{\omega \in \Omega} p(\omega)$ is the *probability* of the event A. It is agreed the $\Pr(\emptyset) = 0$. Clearly $\Pr(\Omega) = 1$.
Example. Assume that $\Omega = \{\omega_1, \dots, \omega_n\}$ is a finite space. Let $p(\omega) = \frac{1}{n}$ for any $\omega \in \Omega$. Then $\Pr(A) = \frac{\#A}{\#\Omega}$. (Note $\#A = 0 \iff A = \emptyset$.) Such probability is called *uniform* distribution.

Let $A, B \in 2^\Omega$ be two events. Then $A \cap B$ is the *intersection* of A and B, is the set which consists of all elements which belong to A and B. $A \cup B$ is the *union* of A and B, is the set which consists of all elements which belong either to A and to B. Then

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B).$$

$A^c := \Omega \backslash A$ is the *complement* of A in Ω, which consists of all points in Ω which are not in A. So

$$A \cup A^c = \Omega,\ A \cap A^c = \emptyset \Rightarrow 1 = \Pr(A) + \Pr(A^c).$$

Let $A_i \subset \Omega, i \in \mathcal{I}$, be a family of subsets of Ω. The the de Morgan rule states $(\cup_{i \in \mathcal{I}} A_i)^c = \cap_{i \in \mathcal{I}} A_i^c$. The sets $A_i, i \in \mathcal{I}$ are called *pairwise disjoint* if $A_i \cap A_j$ for any $i \neq j, i, j \in \mathcal{I}$. Assume that $\mathcal{I}$ is a countable set. Then for any pairwise disjoint countable events $A_i, i \in \mathcal{I}$ in Ω one has

$$\Pr(\cup_{i \in \mathcal{I}} A_i) = \sum_{i \in \mathcal{I}} \Pr(A_i).$$

(This result holds also in the case Ω is not countable sample space.)

Let $A, B \subset \Omega$. Then $A|B$ is the *conditional event* that A will occur if B already occurred. This is equivalent to the event $A \cap B|B$. Assume that $P(B) > 0$. Then $\Pr(A|B)$ is the *conditional probability* of the conditional event $A|B$. Thus $\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$.

1.2 One random variable

$X : \Omega \to \mathbb{R}$ is called a *random variable*. Here $\mathbb{R}$ is the set of real numbers and X is a map from the sample space to $\mathbb{R}$. That is X attaches to each $\omega \in \Omega$ a real number $X(\omega)$. The reason we call X a random variable is follows.

Suppose that $\Omega = \{H, T\}$, which describes the outcome of a coin toss. If the coin falls such that we see head then the outcome is H. If tails is up then the outcome is T. Assume that $X(H) = 1$ and $X(T) = -1$. That is you are paid \$1 if head show up and you lose \$1 if tail shows up. Hence X is a random variable since the outcome of the coin toss is unknown until the coined is tossed . If the coin is *fair* then $\Pr(X = 1) = 0.5$ and $\Pr(X = -1) = 0.5$.

Sometimes it would be convenient to consider more general random variable $X : \Omega \to \Theta$, where Θ is some set, not necessary the set of real numbers. In what follows we shall consider the case $X : \Omega \to \mathbb{R}$.

All the information about the random variable $X : \Omega \to \mathbb{R}$ is stored in the *cumulative distribution function* c.d.f. $F_X : \mathbb{R} \to [0, 1]$:

$$F_X(t) = \Pr(X \le t) = \sum_{X(\omega) \le t} p(\omega).$$

$F_X(t)$ is an increasing, (sometimes called also nondecreasing), i.e. $F_X(t_1) \le F_X(t_2)$ for any $t_1 \le t_2$. Moreover

$$\lim_{t \to -\infty} F_X(t) = 0, \ \lim_{t \to \infty} F_X(t) = 1, \ \lim_{t \searrow x} F_X(t) = x \ \texttt{for all} \ x \in \mathbb{R}.$$

Here $t \searrow x$ stand for t approaches to x from the *right*, i.e. $t > x$ and t converges to x. The last condition of the above conditions means that F_X is continuous form the right.

If $f : \mathbb{R} \to \mathbb{R}$ is a function, then $Y = f(X)$ is the random variable $Y : \Omega \to \mathbb{R}$ such that $Y(\omega) := f(X(\omega))$. The *expected value* of X is denoted by $\mathrm{E}(X)$:

$$\mathrm{E}(X) := \sum_{\omega \in \Omega} p(\omega) X(\omega).$$

Thus if we view each point $\omega \in \Omega$ as a bead of mass $p(\omega)$ concentrated at the point $X(\omega)$ on the real line then $\mathrm{E}(X)$ is the center of mass of the all the beads. If Ω is finite then $\mathrm{E}(X)$ is well defined. If Ω is infinite countable then $\mathrm{E}(X)$ exists if

$$\mathrm{E}(|X|) := \sum_{\omega \in \Omega} p(\omega) |X(\omega)| < \infty.$$

This is true if $|X| \le M \iff |X(\omega)| \le M$ for all $\omega \in \Omega$, for some $M > 0$. The $k - th$ moment of X is defined as $\tau_k := \mathrm{E}(X^k)$ for $k \in \mathbb{N}$. (Note that $\tau_0 := \mathrm{E}(X^0) = \mathrm{E}(1) = 1$.) If Ω is finite the $k - th$ moment exists always. If Ω is infinite countable then $k - th$ moment exists iff $\mathrm{E}(|X|^k) < \infty$. Assume that all the moments of X exist. Then under very mild conditions the moments of X, i.e. the values $\tau_1, \dots$ determine d.c.f. F_X. For example, this is true if $|X| \le M$ for some $M > 0$. More generally it is true if there exist $M > 0$ such that

$$\sum_{i=1}^{\infty} \frac{\mathrm{E}(X^{2k})}{M^{2k}(2k)!} < \infty.$$

The reason that it is enough to consider the even moments is as follows. Recall the Cauchy-Schwarz inequality

$$|\sum_{i=1}^{n} x_i y_i| \le \sum_{i=1}^{n} |x_i| \, |y_i| \,) \le \sqrt{\sum_{i=1}^{n} x_i^2} \, \sqrt{\sum_{i=1}^{n} y_i^2}, \ \texttt{for any} \ x_i, y_i \in \mathbb{R}, \ i = 1, \dots, n. \quad (1.1)$$

Equality holds if there exist $a, b \in \mathbb{R}, a^2 + b^2 > 0$ such that $ax_i = by_i$ for $i = 1, \ldots, n$. ($\mathbf{x} := (x_1, \ldots, x_n)^\top, \mathbf{y} := (y_1, \ldots, y_n) \in^\top \in \mathbb{R}^n$ are *colinear.*) Then for any k we have the inequality

$$|\mathrm{E}(X^k)| \le \mathrm{E}(|X|^k) = \sum_{\omega \in \Omega} \sqrt{p(\omega)}(\sqrt{p(\omega)}|X(\omega)|^k) \le$$
$$\sqrt{\sum_{\omega \in \Omega} p(\omega)} \sqrt{\sum_{\omega \in \Omega} p(\omega) X(\omega)^{2k}} = 1 \cdot \sqrt{\mathrm{E}(X^{2k})}.$$

Hence $\tau_{2k-1}^2 \le \tau_{2(2k-1)}$ for $k \in \mathbb{N}$.

The *variance* of V is defined as the second moment of $X - \mathrm{E}(X)$:

$$\mathrm{Var}(X) := E((X - \mathrm{E}(X))^2) = \mathrm{E}(X^2) - \mathrm{E}(X)^2.$$

Note that the random variable X takes only one value ($\mathrm{E}(X)$) iff $\mathrm{Var}(X) = 0$. Clearly

$$\mathrm{E}(cX) = c\mathrm{E}(X), \quad \mathrm{Var}(cX) = c^2\mathrm{Var}(X) \quad \texttt{for any } c \in \mathbb{R}.$$

Recall that the *standard deviation* of X, denoted by σ_X, is defines as $\sigma_X := \sqrt{\mathrm{Var}(X)}$.

Theorem 1.1 *(Chebyshev's inequality). Let X be random variable with finite $\mathrm{E}(X)$ and $\mathrm{Var}(X)$. Then for any $t > 0$*

$$\Pr(|X - \mathrm{E}(X)| \ge t) \le \frac{\mathrm{Var}(X)}{t^2}.$$

Proof.

$$\mathrm{Var}(X) = \mathrm{E}((X - \mathrm{E}(X))^2) = \sum_{\omega \in \Omega} (X(\omega) - \mathrm{E}(X))^2 = \sum_{\omega \in \Omega, |X(\omega) - \mathrm{E}(X)| \ge t} (X(\omega) - \mathrm{E}(X))^2$$
$$+ \sum_{\omega \in \Omega, |X(\omega) - \mathrm{E}(X)| < t} (X(\omega) - \mathrm{E}(X))^2 \ge \sum_{\omega \in \Omega, |X(\omega) - \mathrm{E}(X)| \ge t} t^2 = \Pr(|X - \mathrm{E}(X)| \ge t) t^2,$$

which implies Chebyshev's inequality. □

X is called *Bernoulli* if $X : \Omega \to \{0, 1\}$, i.e. X takes either value 0 or 1. Assume that X is Bernoulli. Let $\Pr(X = 1) = p \in [0, 1]$. Then $\Pr(X = 0) = 1 - p$. Note that for any $k \in \mathbb{N}$ $X^k = X$. Thus

$$\mathrm{E}(X) = \Pr(X = 0) \cdot 0 + \Pr(X = 1) \cdot 1 = \Pr(X = 1) = p, \quad \mathrm{E}(X^2) = \mathrm{E}(X) = p,$$
$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = p(1 - p). \tag{1.2}$$

To any event $A \subset \Omega$ one associates the following Bernoulli (characteristic) random variable $X_A : \Omega \to \mathbb{R}$: $X_A(\omega) = 1 \iff \omega \in A$. Then $\mathrm{E}(X_A) = \Pr(A)$.

Remark. Assume that Ω is a *general* sample space, i.e. not necessary countable. Then a random variable $X : \Omega \to \mathbb{R}$, (which now satisfies the condition that is a *measurable* map), is called *countable* if $X(\Omega)$ is a countable set in $\mathbb{R}$. Then the treatment of X is equivalent to the treatment of a random variable on a countable space. This is done identifying all points in Ω whose image under X is identical. Then we obtain a countable space Ω' and $X' : \Omega' \to \mathbb{R}$ such that $F_X = F_{X'}$.

1.3 Several random variables

Let $X, Y : \Omega \to \mathbb{R}$ be two random variables. Then

$$\mathrm{Cov}(X, Y) := \mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))) =$$
$$\mathrm{E}(XY) - \mathrm{E}(X\mathrm{E}(Y)) - \mathrm{E}(\mathrm{E}(X)Y) + \mathrm{E}(\mathrm{E}(X)\mathrm{E}(Y)) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) \tag{1.3}$$

is the *covariance* of X and Y. Note that $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$. Apply the Cauchy-Schwarz inequality (1.1) to $|E(XY)|$ to deduce that $E(XY)^2 \le \mathrm{E}(X^2)\mathrm{E}(Y^2)$. Replace X and Y by $X - \mathrm{E}(X)$ and $Y - \mathrm{E}(Y)$ respectively to obtain the inequality

$$\mathrm{Cov}(X,Y)^2 \le \mathrm{Var}(X)\mathrm{Var}(Y).$$

Thus if $\mathrm{Var}(X), \mathrm{Var}(Y) > 0$ we get that $\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \in [-1,1]$. Then

$$\theta := \arccos \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \in [0,\pi]$$

is called the *angle* between X and Y. In particular X and Y are called *orthogonal* if $\mathrm{Cov}(X,Y) = 0$.

Recall that $X+Y : \Omega \to \mathbb{R}$ is the random variable such that $(X+Y)(\omega) = X(\omega)+Y(\omega)$. Clearly $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$. Furthermore

$$\begin{aligned}
&\mathrm{Var}(X+Y) = \mathrm{E}((X+Y-\mathrm{E}(X+Y))^2) = \\
&\mathrm{E}((X-\mathrm{E}(X))^2 + (Y-\mathrm{E}(Y))^2 + 2(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))) \Rightarrow \\
&\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y). \qquad (1.4)
\end{aligned}$$

$X \le Y$ if $X(\omega) \le Y(\omega)$ for all $\omega \in \Omega$. Then

$$X \le Y \Rightarrow \mathrm{E}(X) = \sum_{\omega\in\Omega} p(\omega)X(\omega) \le \sum_{\omega\in\Omega} p(\omega)Y(\omega) = \mathrm{E}(Y). \qquad (1.5)$$

Similarly, if $X_1, \ldots, X_m : \Omega \to \mathbb{R}$ are m random variables then

$$\mathrm{E}(\sum_{i=1}^m X_i) = \sum_{i=1}^m \mathrm{E}(X_i), \quad \mathrm{Var}(\sum_{i=1}^m X_i) = \sum_{i=1}^m \mathrm{Var}(X_i) + \sum_{i\ne j} \mathrm{Cov}(X_i, X_j). \qquad (1.6)$$

Let $X, Y : \Omega \to \mathbb{R}$ be two countable random variables. Then X, Y are called *independent* if

$$\Pr(X = a, Y = b) = \Pr(X = a)\Pr(X = b) \quad \texttt{for all } a, b \in \mathbb{R}.$$

That is the *outcome* of the event $X = a$ is independent of the outcome of the event $Y = b$. Assume that X, Y countable independent random variables. So $X(\Omega) = \{x_i \in \mathbb{R},\ i \in \mathcal{I}\}$ and $Y(\Omega) = \{y_i \in \mathbb{R},\ i \in \mathcal{I}\}$. and Ω is countable. Then

$$\begin{aligned}
&E(XY) = \sum_{i,j\in\mathcal{I}} \Pr(X = x_i, Y = y_j)x_i y_j = \sum_{i,j\in\mathcal{I}} \Pr(X = x_i)\Pr(Y = y_j)x_i y_j \\
&\big(\sum_{i\in\mathcal{I}} \Pr(X = x_i)x_i\big)\big(\sum_{j\in\mathcal{I}} \Pr(Y = y_j)y_j\big) = \mathrm{E}(X)\mathrm{E}(Y).
\end{aligned}$$

In particular $\mathrm{Cov}(X,Y) = 0$. That is two independent random variables are *orthogonal*. Note that for any $x, y \in \mathbb{R}$ $X - x$ and $Y - y$ are also independent.

$X_1, \ldots, X_m : \Omega \to \mathbb{R}$ are called *independent* random variables over a countable sample space Ω if

$$\Pr(X_1 = a_1, X_2 = a_2, \ldots, X_m = a_m) = \Pr(X_1 = a_1) \cdot \Pr(X_2 = a_2) \cdots \Pr(X_m = a_m)$$

for any $a_1, a_2, \ldots, a_m \in \mathbb{R}$. Assume that $X_1, \ldots, X_m$ are independent random variables. Then any subset $X_{i_1}, \ldots, X_{i_l}$, where $1 \le i_1 < \ldots < i_l \le m$, is a set of independent random variables. The arguments for the case $m = 2$ yield

$$\mathrm{E}(X_1 X_2 \cdots X_m) = \mathrm{E}(X_1)\mathrm{E}(X_2) \cdots \mathrm{E}(X_m). \qquad (1.7)$$

Let $X_1, \ldots, X_m$ be random variables such that any pair of random variables X_i, X_j are independent for $i \ne j$. Hence $\mathrm{Cov}(X_i, X_j) = 0$ for any $i \ne j$. The second equality of (1.6) yields

$$\mathrm{Var}(X_1 + X_2 + \ldots + X_m) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \ldots + \mathrm{Var}(X_m). \qquad (1.8)$$

The above equality holds if $X_1, \ldots, X_m$ are independent random variable.

1.4 Inclusion-Exclusion Principle

The following fact about Bernoulli random variables is straightforward and its proof is left to the reader.

Proposition 1.2 *Let Ω be a sample space and $X_1, \ldots, X_k : \Omega \to \{0,1\}$ be Bernoulli random variables. Let $A_i := \{\omega \in \Omega : X_i(\omega) = 1\}$ for $i = 1, \ldots, k$. Then $X_i = X_{A_i}$ and $1 - X_i = X_{A_i^c}$ is Bernoulli for $i = 1, \ldots, k$. Furthermore $X = X_1 \cdot X_2 \cdots X_k$ is Bernoulli and $X = X_{A_1 \cap A_2 \cap \ldots \cap A_k}$. In particular $\mathrm{E}(X_1 \cdot X_2 \cdots X_k) = \Pr(A_1 \cap A_2 \cap \ldots \cap A_k)$.*

Lemma 1.3 *Let $x_1, \ldots, x_n \in \mathbb{C}$. Then*

$$(1 - x_1) \cdot (1 - x_2) \cdots (1 - x_n) = 1 + \sum_{k=1}^{n} (-1)^k \sum_{1 \le i_1 < \ldots < i_k \le n} x_{i_1} \cdots x_{i_k}. \tag{1.9}$$

Assume furthermore that $x_1, \ldots, x_n \in \{0,1\}$. Then for any even integer $2p \in [0,n]$ and odd integer $2q-1 \in [1,n]$ one has the inequalities

$$1 + \sum_{k=1}^{2q-1} (-1)^k \sum_{1 \le i_1 < \ldots < i_k \le n} x_{i_1} \cdots x_{i_k} \le \prod_{i=1}^{n} (1 - x_i) \le 1 + \sum_{k=1}^{2p} (-1)^k \sum_{1 \le i_1 < \ldots < i_k \le n} x_{i_1} \cdots x_{i_k}. \tag{1.10}$$

Proof. Equality (1.9) is straightforward and can be proven by induction on n. The inequalities (1.10) are proved as follows. Assume that m out n variables $x_1, \ldots, x_n$ are equal to 1. If $m = 0$ then $x_1 = \ldots = x_n = 0$ and we have all the expression in (1.10) are equal to 1. Hence (1.10) holds. Assume that $m \in [1,n]$. Without loss of generality we may assume that $x_1 = \ldots = x_m = 1$ and $x_{m+1} = \ldots x_n = 0$. In that case $(1-x_1) \cdots (1-x_n) = 0$. Observe next that $x_{i_1} \ldots x_{i_l} = 0$ for any $1 \le i_1 < \ldots < i_l \le n$ and $l > m$. It the follows that

$$\sum_{1 \le i_1 < \ldots i_l \le n} x_{i_1} \cdots x_{i_l} = \binom{m}{l} \quad \texttt{for any integer } l \in [1,n].$$

Indeed, this equality corresponds to choose l $x_{i_1}, \ldots, x_{i_l}$ out of $\{x_1, \ldots, x_m\}$ which are al equal to 1. Thus (1.10) is equivalent to

$$\sum_{k=0}^{2q-1} (-1)^k \binom{m}{k} \le 0 \le \sum_{k=0}^{2p} (-1)^k \binom{m}{k} \tag{1.11}$$

for any $p \in \mathbb{Z}_+$ and $q \in \mathbb{N}$. Recall that the sequence $\binom{m}{l}$ is nondecreasing for $l = 0, 1, \ldots, \lceil \frac{m}{2} \rceil$. Since

$$\sum_{k=0}^{2q-1} (-1)^k \binom{m}{k} = \sum_{k=0}^{q-1} \binom{m}{2k} - \binom{m}{2k+1}$$

it follows that for $2q - 1 \le \lceil \frac{m}{2} \rceil$ we deduce the first inequality in (1.11). Since $\binom{m}{0} = 1$ we clearly have the second inequality in (1.11) for $p = 0$. For $p \ge 1$ we have the identity

$$\sum_{k=0}^{2p} (-1)^k \binom{m}{k} = 1 + \sum_{k=1}^{2p} \binom{m}{2k} - \binom{m}{2k-1}.$$

Hence for $2p \le \lceil \frac{m}{2} \rceil$ we deduce the second inequality in (1.11).

As $\binom{m}{l} = 0$ for $l > m$ it is enough to prove (1.11) for $m \ge 2q-1, 2p \ge \lceil \frac{m}{2} \rceil$. Recall that $0 = (1-1)^m = \sum_{k=0}^{m} (-1)^k \binom{m}{k}$. Subtract this identity from both sides of (1.11) and use the identities $\binom{m}{k} = \binom{m}{m-k}$ for $k = 1, \ldots, m$ to deduce the cases $m \ge 2q-1, 2p \ge \lceil \frac{m}{2} \rceil$ from the cases $2q-1, 2p \le \lceil \frac{m}{2} \rceil$. □

Theorem 1.4 *Let $A_1, \ldots, A_n \subset \Omega$ be n events in a sample space Ω. Then*

$$\Pr(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i=1}^{n} \Pr(A_i) + \sum_{k=2}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \ldots < i_k \le n} \Pr(A_{i_1} \cap \ldots \cap A_{i_k}). \quad (1.12)$$

Furthermore for any even integer $2p \in [1, n]$ and odd integer $2q-1 \in [1, n]$ one has

$$\sum_{k=1}^{2p} (-1)^{k-1} \sum_{1 \le i_1 < \ldots < i_k \le n} \Pr(A_{i_1} \cap \ldots \cap A_{i_k}) \le \Pr(\cup_{i=1}^{n} A_i) \le$$
$$\sum_{k=1}^{2q-1} (-1)^{k-1} \sum_{1 \le i_1 < \ldots < i_k \le n} \Pr(A_{i_1} \cap \ldots \cap A_{i_k}). \quad (1.13)$$

Proof. Let $X_i = X_{A_i}, Y_i = X_{A_i^c} = 1 - X_i, i = 1, \ldots, n$. Then

$$\Pr(\cup_{i=1}^{n} A_i) = 1 - \Pr((\cup_{i=1}^{n} A_i)^c) = 1 - \Pr(\cap_{i=1}^{n} A_i^c) = 1 - \mathrm{E}(Y_1 \cdots Y_n) = 1 - \mathrm{E}((1-X_1)\cdots(1-X_n)).$$

Use expansion (1.9) and Proposition 1.2 to deduce

$$\mathrm{E}((1-X_1)\cdots(1-X_n)) = \mathrm{E}(1 + \sum_{k=1}^{n} (-1)^k \sum_{1 \le i_1 < \ldots i_k \le n} X_{i_1} \cdots X_{i_k}) =$$
$$1 + \sum_{k=1}^{n} (-1)^k \sum_{1 \le i_1 < \ldots i_k \le n} \mathrm{E}(X_{i_1} \cdots X_{i_k}) = 1 + \sum_{k=1}^{n} (-1)^k \sum_{1 \le i_1 < \ldots i_k \le n} \Pr(A_{i_1} \cap \ldots \cap A_{i_k}).$$

Combine the above two equalities to deduce (1.12). Since each $X_i \in \{0, 1\}$ we can apply inequality (1.10) to deduce

$$1 + \sum_{k=1}^{2q-1} (-1)^k \sum_{1 \le i_1 < \ldots < i_k \le n} X_{i_1} \cdots X_{i_k} \le \prod_{i=1}^{n} (1 - X_i) \le 1 + \sum_{k=1}^{2p} (-1)^k \sum_{1 \le i_1 < \ldots < i_k \le n} X_{i_1} \cdots X_{i_k}.$$

Take the expected value of all the three random variables appearing in the above inequality, use (1.5) and the above arguments to obtain (1.13). □

Remark. The equality (1.12) is called the *inclusion-exclusion principle.* The inequalities (1.13) are called the *Bonferroni* inequalities.

1.5 Binomial and Poisson random variables

Random variables $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ are called be i.d., (identically distributed random variables), if $F_{X_1} = \ldots = F_{X_n}$. $X_1, \ldots, X_n$ are called i.i.d., (independent,identically distributed random variables), in addition to i.d. these variables are independent.

Let $X_1, \ldots, X_n : \Omega \to \{0, 1\}$ be Bernoulli. Define $Y = X_1 + \ldots + X_n$ then $X : \Omega \to \{0, 1, \ldots, n\}$ have nonnegative integer values in $[0, n]$. The exact distribution of X depends on the *joint* distribution of $X_1, \ldots, X_n$. For each integer $k \in [0, n]$ $\Pr(X = k)$ can be expressed as the expectation of the following Bernoulli random variable W_k for $k = 0, 1, \ldots, n$. Let $W_0 = (1 - X_1) \cdots (1 - X_n)$. Then $\Pr(Y = 0) = \mathrm{E}(W_0)$. Consider the random variable $U_k = X_1 \ldots X_k (1 - X_{k+1}) \cdots (1 - X_n)$. Then U_k is Bernoulli with

$$U_k = 1 \iff X_1 = \ldots = X_k = 1,\ X_{k+1} = \ldots = X_n = 0.$$

Hence

$$\Pr(X_1 + \ldots + X_n = k) = \mathrm{E}(W_k),\ W_k = \sum_{1 \le i_1 < \ldots < i_k \le n} X_{i_1} \cdots X_{i_k} \prod_{j \ne i_1, \ldots, j \ne i_k} (1 - X_j), \quad (1.14)$$

for $k = 0, 1, \ldots, n$.

In the case $X_1, \ldots, X_n$ are i.i.d. Bernoulli one can find the distribution of $Y := Y(p, n)$ using one parameter $p = \mathrm{E}(X_1) = \ldots = \mathrm{E}(X_n)$. Indeed $Y = k$ if exactly $X_{i_1} = \ldots = X_{i_k} = 1, 1 \le i_1 < \ldots < i_k \le n$, with probability p^k, while all the other variables take the value 0, with probability $(1-p)^{n-k}$. Hence the probability of the above event is $p^k(1-p)^{n-k}$. Since we can choose $1 \le i_1 < \ldots < i_k \le n$ in $\binom{n}{k}$ ways it follows

$$\Pr(Y(n,p) = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n. \tag{1.15}$$

The random variable $Y(n, p)$ is called *binomial* with parameters n and p. Note that from the definition of $Y(n, p)$ as sum of n Bernoulli it follows $\mathrm{E}(Y(n, p)) = np$.

Let $X : \Omega \to \mathbb{Z}_+$ be a countable random variable. Then $\mathrm{Pu}(a)+ : X$ is called *Poisson*, if X has the following distribution:

$$\Pr(\mathrm{Pu}(a) = k) = e^{-a} \frac{a^k}{k!} \quad \texttt{for } k = 0, 1, \ldots, \texttt{ and some } a \ge 0.$$

Hence

$$\mathrm{E}(\mathrm{Pu}(a)) = \sum_{k=0}^{\infty} e^{-a} \frac{a^k}{k!} k = a \sum_{k=1} e^{-a} \frac{a^{k-1}}{(k-1)!} = a.$$

It is possible to obtain $\mathrm{Pu}(a)$ as limit of the binomial binomial with certain parameters.

Proposition 1.5 *Let $Y(n, p_n) = X_{1,n} + \ldots + X_{n,n}$, where $X_{1,n}, \ldots, X_{n,n}$ are i.i.d. Bernoulli with $\mathrm{E}(X_{1,n}) = p_n \in [0, 1]$ for $n \in \mathbb{N}$. Assume that there exists a subsequence $1 \le n_1 < n_2 < \ldots$ such that $\lim_{m\to\infty} \mathrm{E}(Y(n_m, p_{n_m})) = \lim_{m\to\infty} n_m p_{n_m} = a$. Then $X_{n_1}, X_{n_2}, \ldots$ converge in probability to $\mathrm{Pu}(a)$. That is*

$$\lim_{m\to\infty} \Pr(Y(n_m, p_{n_m}) = k) = e^{-a} \frac{a^k}{k!} \quad \texttt{for } k = 0, 1, \ldots.$$

Proof. Recall that

$$\Pr(Y(n,p) = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{1}{k!}(1 - \frac{1}{n}) \ldots (1 - \frac{k-1}{n})(np)^k (1-p)^n (1-p)^{-k}.$$

Now

$$\lim_{m\to\infty} (n_m p_{n_m})^k = a^k, \ \lim_{m\to\infty} (1 - p_{n_m})^{n_m} = e^{-a}, \ \lim_{m\to\infty} (1 - p_{n_m})^{-k} = 1$$

and the proposition follows. □

It is possible to deduce the conclusion of this proposition assuming much less than i.i.d. Bernoulli $X_{1,n}, \ldots, X_{n,n}$.

Theorem 1.6 *Let $1 \le n_1 < n_2 < \ldots$ be a n increasing sequence of integers. Assume that $Z_{1,m}, \ldots, Z_{n_m,m}$ be Bernoulli. Let $Y_m := \sum_{i=1}^{n_m} Z_{i,m}$. Suppose that for each k*

$$\lim_{m\to\infty} \mathrm{E}\Big(\sum_{1 \le i_1 < \ldots < i_k \le n_m} Z_{i_1,m} \cdots Z_{i_k,m} \Big) = \frac{a^k}{k!} \quad \texttt{for } k = 0, 1, \ldots \texttt{ and } a \ge 0. \tag{1.16}$$

Then $Y_1, Y_2, \ldots$ converges in probability to $\mathrm{Pu}(a)$. That is

$$\lim_{m\to\infty} \Pr(Y_m = k) = e^{-a} \frac{a^k}{k!} \quad \texttt{for } k = 0, 1, \ldots. \tag{1.17}$$

Proof. We first prove (1.17) for $k = 0$. Note that

$$Y_m = 0 \iff Z_{1,m} = \ldots = Z_{n_m,m} = 0 \iff \prod_{i=1}^{n_m}(1 - Z_{i,m}) = 1.$$

Hence $\Pr(Y_m = 0) = \mathrm{E}(\prod_{i=1}^{n_m}(1-Z_{i,m}))$. Use the arguments of the proof of (1.13) to deduce that

$$1 + \sum_{k=1}^{2q-1}(-1)^k \sum_{1\le i_1<\ldots<i_k\le n_m} \mathrm{E}(Z_{i_1,m})\cdots Z_{i_k,m}) \le \mathrm{E}(\prod_{i=1}^{n_m}(1 - Z_{i,m})) = \Pr(Y_m = 0) \le$$
$$1 + \sum_{k=1}^{2p}(-1)^k \sum_{1\le i_1<\ldots<i_k\le n_m} \mathrm{E}(Z_{i_1,m}\cdots Z_{i_k,m}).$$

Let $m \to \infty$ and use the assumption (1.16) to deduce

$$1 + \sum_{k=1}^{2q-1}(-1)^k \frac{a^k}{k!} \le \liminf_{m\to\infty} \mathrm{E}(Y_m) \le \limsup_{m\to\infty} \mathrm{E}(Y_m) \le 1 + \sum_{k=1}^{2p}(-1)^k \frac{a^k}{k!}.$$

Recall that $e^{-a} = \lim_{l\to\infty} 1 + \sum_{k=1}^{l}(-1)^k \frac{a^k}{k!}$. Let $p, q \to \infty$ in the above inequalities to deduce (1.17) for $k = 0$.

To prove (1.17) one need to use (1.14). Let

$$W_{k,m} = \sum_{1\le j_1<\ldots<j_k\le n_m} Z_{j_1,m}\cdots Z_{j_k,m} \prod_{j\ne j_1,\ldots,j\ne j_k} (1 - Z_{j,m}).$$

For each $\prod_{j\ne j_1,\ldots,j\ne j_k}(1 - Z_{j,m})$ use the inequalities (1.10) with fixed p and q. This will give lower and upper bounds for $\mathrm{E}(W_{k,m})$. Let $m \to \infty$ and use (1.16) to obtain lower and upper bounds on $\liminf_{m\to\infty} \mathrm{E}(W_{k,m}) \le \limsup_{m\to\infty} \mathrm{E}(W_{k,m})$ as in the case $k = 0$. Now let $p, q \to \infty$ to deduce (1.17) for any $k \ge 1$. □

Remark. We note that the assumptions of Proposition 1.5 imply the conditions of Theorem 1.17. Indeed $\mathrm{E}(X_{i_1,n}\cdots X_{i_k,n}) = p_n^k$ for any $1 \le i_1 < \ldots < i_k \le n$. Hence $\mathrm{E}(\sum_{1\le i_1<\ldots<i_k\le n} X_{i_1,n}\cdots X_{i_k,n}) = \binom{n}{k}p_n^k$. Let $Z_{i,m} = X_{i,n_m}$ for $i = 1, \ldots, n_m$. The assumption that $\lim_{m\to\infty} n_m p_{n_m} = a$ yields (1.16) for $k \in \mathbb{Z}_+$.

2 Graphs

2.1 Undirected Graphs 1-12-05

An *undirected* graph $G := (V, E)$ consists of *finite* sets of vertices V and edges E. Each edge is an *unordered* pair of $(a, b)(= ab)$ of two distinct vertices $a \ne b \in V$. Sometimes we let $V = V(G), E = E(G)$ to emphasize that V and E correspond to the graph G. In this section all the graphs assumed to be *undirected*.

The *cardinality* $\#V$ is called the the *order* of G. Let $n = \#V$. Then it would be convenient to identify $V = \langle n\rangle := \{1, \ldots, n\}$. The *cardinality* $\#E$ is called the *size* of G.

Two graphs $G = (V, E)$ and $H = (W, F)$, having the same order and size, are called *isomorphic* if there exists a *bijection* $\phi : V \to W$ such that $(u, v) \in E \iff (\phi(u), \phi(v)) \in F$. $G \cong H$ denotes that G and H are isomorphic.

A *complete n-graph* is the graph on n vertices denoted by $K_n := (\langle n\rangle, E^n)$, were E^n is the set $\binom{n}{2} := \frac{n(n-1)}{2}$ of all edges (i, j) for $i = 1, \ldots, n, j = i + 1, \ldots, n$. Any graph $G = (\langle n\rangle, E)$ on n vertices is a subgraph of K_n. The complement of G is $G^c := (\langle n\rangle, E^n\backslash E)$. (For any two subsets P, Q of a given set R, $P\backslash Q$ stands for all elements in P which are not

in Q. Note that $P \backslash Q$ may be an empty set $\emptyset$.) The *empty n-graph* $K_n^c := (\langle n \rangle, \emptyset)$ is the complement of K_n. $K_1 = K_1^c$ is called *trivial* graph or *isolated* vertex.

Two vertices $a, b \in V$ are called *adjacent* $\iff (a, b) \in E$. Let $G = (V, E)$ be a graph and $v \in V$. Then $\Gamma(v)$ denotes the *neighborhood* of v, i.e. the set of all neighbors of v. $\deg(v) := \#\Gamma(v)$, i.e. the number of neighbors of v, is called the *degree* of the vertex v. Let $\#V = n$. Then the *degree sequence* of G is the sequence of the degrees of all the vertices of G arranged in a decreasing order: $\deg(v_1) \geq \deg(v_2) \geq \ldots \geq \deg(v_n) \geq 0$. Since every edge in G is connected to two distinct vertices it follows

$$\sum_{v \in V} \deg(v) = 2\#E. \tag{2.1}$$

A *path* P of *length* $l \geq 1$ in G is given the set of vertices $V(P) = \{v_0, v_1, \ldots, v_l\}$ such that $(v_{i-1}, v_i) \in E$ for $i = 1, \ldots, l$ and $v_i \neq v_j$ for $i \neq j$. A *walk* P of *length* $l \geq 1$ in G is given the set of vertices $V(W) = \{v_0, v_1, \ldots, v_l\}$ such that $(v_{i-1}, v_i) \in E$ for $i = 1, \ldots, l$. W is a closed walk if $v_l = v_0$. A walk W is called a *trail* if all the edges of in the walk W $(v_0, v_1), (v_1, v_2), \ldots, (v_{l-1}, v_l)$ are distinct. A closed trail is called a *circuit*. A circuit W is called a *cycle* if $v_i \neq v_j$ for $0 \leq i < j \leq l - 1$. Note that the closed walk $\{i, j.i\}$ for $1 \leq i < j \leq n$ in K_n is not a cycle. We call such a closed walk as *semi-cycle*. Thus any cycle in undirected graph has length 3 at least. The following result is straightforward and its proof is left to the reader.

Proposition 2.1 *Let W be a closed walk on an undirected graph G. Then the edges of W can be decomposed to a union of the edges of cycles and semi-cycles.*

Let $u, v \in V$ be two distinct vertices. Then u is *connected* to v if there exists a path P starting at u and ending at v. We denote $u \sim v$ if u is connected to v. Clearly $u \sim v \iff v \sim u$. Thus $u \sim v$ if u and v are connected. It is convenient to assume that $u \sim u$, i.e. u is connected to itself. Then $\sim$ is an *equivalence* relation on V.

Definition 2.2 *Let V be a set of elements and $\sim$ be a relation on some pairs of the elements of V. That is, there exists a subset $\mathcal{P} \subset V \times V$ such that $u \sim v \iff (u, v) \in \mathcal{P}$.*
(a) $\sim$ is called reflexive if for each $v \in V$ $v \sim v$.
(b) $\sim$ is called symmetric if $u \sim v \iff v \sim u$.
(c) $\sim$ is called transitive if $u \sim v$, $v \sim w \Rightarrow u \sim w$.
$\sim$ is called an equivalence relation if $\sim$ is reflexive, symmetric and transitive.

Proposition 2.3 *Let V be a set with an equivalence relation $\sim$. Then V decomposes to a disjoint union of nonempty equivalence classes $V_i, i \in \mathcal{I}$ such that:*
(a) For any $i \in \mathcal{I}$ and $u, v \in V_i$ $u \sim v$.
(b) For any $i, j \in \mathcal{I}, i \neq j$ and $u \in V_i, v \in V_j$ one has $u \nsim v$.

Corollary 2.4 *Let $G = (V, E)$ be an undirected graph. Then there is a unique decomposition of V to disjoint union of nonempty subsets of vertices $V = \cup_{i=1}^k V_k$, (where the uniqueness is up to a permutation of the $V_1, \ldots, V_k$), such that the following conditions hold:*
(a) $G = \cup_{i=1}^k G(V_i)$.
(b) Any two vertices in each $G(V_i)$ are connected.

$G(V_1), \ldots, G(V_k)$ are called the *connected components* of G.

Definition 2.5 *Let $G = (V, E)$ be an undirected graph. G is called acyclic if G does not have a cycle. An acyclic connected G is called a tree. An acyclic G is called a forest.*
A subgraph $T = (V, W), W \subset E$ of a connected G is called a spanning tree if T is a tree.

Let $G = (V, E)$ be a connected digraph. For $u, v \in G$ we define the *distance* $\mathrm{dist}(u, v)$ as follows:
(a) $\mathrm{dist}(v, v) = 0$ for all $v \in V$.
(b) For $u \neq v \in V$ $\mathrm{dist}(u, v)$ is the shortest length of a path connecting u to v.

Clearly $\mathrm{dist}(u, v) = \mathrm{dist}(v, u)$ (*symmetricity*). Note that if W is a walk of length l from u to v then $l \geq dist(u, v)$. Equality holds iff W is a shortest path from u to v. Hence the distance function satisfies the triangle inequality: $\mathrm{dist}(u, w) \leq \mathrm{dist}(u, v) + \mathrm{dist}(v, w)$ for any $u, v, w \in V$. Thus dist is a function $\mathrm{dist}(\cdot, \cdot) : V \times V \to [0, \infty)$ which satisfies property (a), symmetricity, triangle inequality and $\mathrm{dist}(u, v) > 0$ for any $u \neq v \in V$. Thus $\mathrm{dist}(\cdot, \cdot)$ is the *distance function* on V.

Proposition 2.6 *Let $G = (V, E)$ be a connected undirected graph. Then G has a spanning tree.*

Proof. If G is trivial, i.e. $\#V = 1$, then G is its spanning tree. Assume that $n = \#V > 1$. Let $V_0 := \{v_0\}$. For each $i \in \mathbb{N}$ let $V_i := \{v \in V : \mathrm{dist}(v_0, v) = i\}$. Then there exists positive integer k such that $V_i \neq \emptyset$ for $i = 1, \ldots, k$ and $V_i = \emptyset$ for $i > k$. By the definition $V_i \cap V_j = \emptyset$ for $i \neq j$. Since G is connected $V = \cup_{i=0}^{k} V_i$. Also $E \cup V_i \times V_j = \emptyset$ for $j - i \geq 2$. Now let $E_i = E \cap V_{i-1} \times V_i$ for $i = 1, \ldots, k$. By the definitions of $V_i, i = 0, \ldots, k$ $E_i \neq \emptyset$ for $i = 1, \ldots, k$. Then $T := (V, \cup_{i=0}^{k} E_i)$ is a spanning tree of G. □

The tree described in the proof of the above proposition is called a *rooted* tree, with the root v_0. It is straightforward to show:

Proposition 2.7 *The following assertion are equivalent for an undirected graph $G = (V, E)$:*
(a) G is a tree.
(b) G is a minimal connected graph on V. That is G is connected and for any edge $(u, v) \in E$ the subgraph $H = (V, E \backslash \{(u, v)\})$ is disconnected.
(c) G is a maximal acyclic graph. That is G is acyclic and for any $u \neq v \in V$ such that that $(u, v) \notin E$ the graph $H = (V, E \cup \{(u, v)\})$ contains a cycle.

Proposition 2.8 *A tree of order n has size $n-1$. A forest of order n with k components has size $n - k$.*

Proof. Assume first that $G = (V, E)$ is a tree. We show that $\#E = \#V - 1$ by induction on $n := \#V$. For $n = 1$ $G = K_1$ and $E = \emptyset$. Assume that the claim holds for all trees of order m at most. Let $G = (V, E)$ be a tree with $\#V = m + 1$. Let $(u, v) \in E$ ad consider the subgraph $H = (V, E \backslash (u, v)$. Then G is a union of two disjoint trees $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ each of order m at most. The induction hypothesis yields that $\#E_1 = \#V_1 - 1, \#E_2 = \#V_2 - 1$. Hence $\#E = \#E_1 + \#E_2 + 1 = \#V_1 + \#V_2 - 1 = \#V - 1$.

To prove the corresponding claim for the forest, use the fact that a forest with k components is a disjoint union of k trees. □

Corollary 2.9 *Any tree of size 2 at least has at least two vertices of degree 1.*

Proof. Assume that G is a tree of order $n \geq 2$. Let $d_1 \geq \ldots \geq d_n$ be the degree sequence of G. Since G is connected $d_n \geq 1$. Since the size of G is $n - 1$ it follows that $\sum_{i=1}^{n} d_i = 2(n - 1)$. Since each $d_i \in \mathbb{N}$ it follows that we must have $d_{n-1} = d_n = 1$. Otherwise if $d_{n-1} \geq 2$ and $d_n \geq 1$ we will have that $\sum_{i=1}^{n} d_i \geq 2(n - 1) + 1 > 2(n - 1)$ which contradicts the above equality. □

A vertex in a tree of size 2 at least is called a *leaf* if its degree is 1. A tree of size 2 at least which have exactly two leaves is called a path: $\bullet - \bullet - \ldots - \bullet - \bullet$. A tree of size $n \geq 2$ which has $n - 1$ leaves is called a star. It is isomorphic to $K_{1,n-1} := (\langle n \rangle, \cup_{i=2}^{n}(1, i))$.

$G = (V, E)$ is called *bipartite* if V decomposes to two disjoint nonempty sets V_1, V_2 such that $E \subset V_1 \times V_2$.

Theorem 2.10 *Let G be an indirected graph. Then G is bipartite iff G does not have cycles of odd length.*

Proof. Clearly G is bipartite iff each connected component is bipartite. Thus it is enough to prove the theorem in the case $G = (V, E)$ is connected. If G is bipartite then any walk is of the form $V_1 - V_2 - V_1 \ldots$ or $V_2 - V_1 - V_2 \ldots$. In particular any closed walk on G has an even length. In particular any cycle has an even length.

Assume that any cycle, if exists, has even length. Since the length of a semi-cycle is 2 Proposition 2.1 yields that any closed walk in G has an even length. Hence any two walks between any two vertices $u, v \in V$ have the same parity. Fix $v_0 \in V$. Let V_1 be the set of all vertices in $v \in V$ such that any walk from v_0 to v has an odd length. Then $V_2 := V \backslash V_1$ is the set of all the vertices $v \in V$ such that any walk from v_0 to v has an even length. Hence $E \subset V_1 \times V_2$. □

Let U, V be two disjoint sets of vertices of cardinality m, n respectively. The the complete bipartite graph $K_{m,n}$ has order $m + n$ and size mn, and is obtained by joining every vertex $u \in U$ to every vertex $v \in V$. $K_{m,n}$ is isomorphic to $G := (U \cup V, U \times V)$. Note that star $K_{1,n-1}$ is a complete bipartite graph on $\#U = 1, \#V = n - 1$.

Definition 2.11 *Let $G = (V, E)$ be an undirected connected graph. C is called a Hamiltonian cycle in G if C is a cycle on all vertices of G. P is called a Hamiltonian path in G if P is a path on all vertices of G. C is called Eulerian circuit if C is a circuit of G containing all edges of E. T is called an Eulerian trail if T is trail that contains all the the edges of G.*

Theorem 2.12 *Let $G = (V, E)$ be an undirected connected nontrivial graph. Then G has an Eulerian circuit iff the degree of each vertex in G is even. G contains a non-closed Eulerian trail iff G has two vertices of odd degree and all other vertices have even degrees.*

Proof. When travelling on a Eulerian circuit every time you enter the vertex v you exit it, it follows that the degree of each vertex is even. We prove by induction that on the size $m = \#V$, that if all the vertices of G have even degree than G contains an Eulerian circuit. For $m = 3$ $G = K_3$ and the unique cycle on K_3 is Eulerian. Assume that the theorem holds for any G with of size $m \geq 3$. Assume that $\#E = m + 1$. Then $\#V > 3$. Since each vertex of v has an even degree, G can not be a tree, hence there exists a cycle C in G. Consider a subgraph of $G_1 = (V, E_1)$ obtained from G by deleting all edges in the cycle C. Suppose first that G_1 is connected. Then the induction hypothesis implies that G_1 has a Eulerian circuit C_e. C_e can be started from any vertex $v \in V$. Start it from the vertex v in the cycle C. Complete first C_e ending at v and then complete the cycle C to obtain an Eulerian circuit on G. If G_1 is disconnected then any vertex in any nontrivial connected component H has even degree. The induction hypothesis yields that each nontrivial component H has an Eulerian circuit. It is straightforward to show how to combine these Eulerian circuits with C to obtain a Eulerian circuit on G.

Assume that G contains a non-closed Eulerian trail T. Then the initial and the end vertex T have odd degrees and all other vertices have even degrees. Suppose that a connected G has two vertices u, v of odd degrees and all other vertices of V have even degree. Let P be a path from u to v in G. Let $G_1 = (V, E_1)$ be the a subgraph of G obtained by deleting all edges on the path P. Then each vertex $v \in V$ has even degree in G_1. Hence every nontrivial component of G_1 has an Eulerian circuit. Combine these Eulerian circuits with the path P to obtain a non-closed Eulerian trail from u to v. □

Let $\mathbf{P}$ be a graph property. For example: $\mathbf{P}_E$ - a graph G has an Eulerian circuit; $\mathbf{P}_H$ - a graph G has a Hamiltonian cycle. How "difficult" is to find out if G has the property $\mathbf{P}$? By difficult we mean computational complexity of $\mathbf{P}$, i.e. how many computational

operations we need in the *worst case scenario* to find if a given G has property $\mathbf{P}$. Clearly the amount of computation depend on the size n of G.

Definition 2.13 *A property* $\mathbf{P}$ *of a graph* G *is called polynomial if there exists a polynomial* $p(x) = a_0x^l + a_1x^{l-1} + \ldots + a_l$ *such that for every* $G = (\langle n \rangle, E)$ *one needs at most* $p(n)$ *operation to find out if* G *has property* $\mathbf{P}$ *for any* $n \in \mathbb{N}$. *The set of all polynomial properties* $\mathbf{P}$ *of graphs is denoted by* $\mathbf{Po}$.

A property $\mathbf{P}$ *is called nondeterministic polynomial if there exists a polynomial* $q(x) = b_0x^m + b_1x^{m-1} + \ldots + b_m$ *with the following property. For every* $G = (\langle n \rangle, E)$ *that is claimed by somebody to have the property* $\mathbf{P}$ *with the provided documentation of the claim, it will take you at most* $q(n)$ *steps to find if the person is right. The set of all nondeterministic polynomial* $\mathbf{P}$ *is denoted by* $\mathbf{NP}$.

The property $\mathbf{P}_E$ is a polynomial property in view of Theorem 2.12. The property $\mathbf{P}_H$ is nondeterministic polynomial. Indeed to substantiate a claim that a given graph $G = (\langle n \rangle, E)$ has a Hamiltonian cycle one should supply the Hamiltonian cycle C. Now it is your task to check out if C is indeed a Hamiltonian cycle in G as the person claims. This check is clearly polynomial in n.

It is straightforward to show that $\mathbf{Po} \subseteqq \mathbf{NP}$. It is not known if $\mathbf{Po} \subsetneqq \mathbf{NP}$. (There is a one million dollar prize for the solution of this problem!) It is believed that $\mathbf{NP}$ is strictly larger than $\mathbf{Po}$. It is known that $\mathbf{NP}$ contains a subset of the most difficult problems called $\mathbf{NP}$-*complete*, which are denoted by $\mathbf{NPC}$. Every two problems in the class $\mathbf{NPC}$ are polynomially equivalent. Put it differently, if $\mathbf{P} \in \mathbf{NPC}$ is polynomial the $\mathbf{Po} = \mathbf{NPC}$. Roughly speaking any problem in $\mathbf{NPC}$ is believed to have an exponential complexity. It is known that $\mathbf{P}_H \in \mathbf{NPC}$.

2.2 Directed Graphs 2-4-05

A *directed* graph, sometimes referred as *digraph*, is $G := (V, E)$, where V is the set of vertices and $E \subset V \times V$ is the set of edges. In what follows we assume that $G = (V, E)$ a digraph, unless stated otherwise. An edge of the form $(u, v) \in V$, where $u \neq v$ is called the edge from u to v. An edge of the form $(u, u) \in E$ for some $u \in G$ is called a *loop* (on u). A digraph without loops is called *loopless*. A *path* P in G is given by the vertices $V(P) = \{v_0, v_1, \ldots, v_l\}$ where $v_i \neq v_j$ for $i \neq j$ and $(v_{i-1}, v_i) \in E$ for $i = 1, \ldots, l$. A *trail* T in G is given by the vertices $V(T) = \{v_0, v_1, \ldots, v_l\}$, where $(v_{i-1}, v_i) \in E, i = 1, \ldots, l$ is are l distinct edges in G. A *cycle* C of length l is given by a trail T as above, such that $v_l = v_0$ and $v_i \neq v_j$ for $1 \leq i < j \leq l$. A cycle of length 1 is a loop on v_0 and a cycle of length 2 is given by the two edges $(v_0, v_1), (v_1, v_0)$. A *walk* W of length $l \geq 1$ is given by the $V(W) = \{v_0, v_1, \ldots, v_l\}$, where $(v_{i-1}, v_i) \in E, i = 1, \ldots, l$.

For $v \in V$ let

$$\deg_{\text{out}}(v) := \#\{u \in V : \ (v, u) \in E\}, \quad \deg_{\text{in}}(v) := \#\{u \in V : \ (u, v) \in E\},$$

be the *out* and *in* degree of the vertex v. Note that if the loop $(v, v) \in E$, then it contributes to out and in degree of v. It is straightforward to show that

$$\sum_{v \in V} \deg_{\text{out}}(v) = \sum_{v \in V} = \deg_{\text{in}}(v) = \#E.$$

If $H = (V, E_{\text{undir}})$ is an undirected graph, there are two standard ways to associate with H a digraph $G = (V, E)$ on the same set of vertices V. The maximal directed graph $G = (V, E_{\max})$ of H is given by letting (u, v) and (v, u) be in $E_{\max}$ if and only if the indirected edge uv is in E_{undir}. A minimal directed graph $G = (V, E_{\text{orient}}), E_{\text{orient}} \subset E_{max}$, such that if $uv \in E_{\text{undir}}$ then either $(u, v) \in E_{\text{orient}}$ or $(v, u) \in E_{\text{orient}}$ but not both. This is equivalent to an orientation of any undirected edge $uv \in E_{\text{undir}}$. Vice versa, any digraph

$G = (V, E)$ induces the following unique undirected graph $H = (V, E_{\text{undir}})$ on the same set of vertices: For $u, v \in V$ $uv \in E_{\text{undir}}$ iff $u \neq v$ and either (u, v) or (v, u) are in E, (or both).

$G_1 = (V_1, E_1)$ is a subgraph of a digraph G if $V_1 \subset V, E_1 \subset E$. Let $V_1 \subset V$. Then $G(V_1) := (V_1, E \cap (V_1 \times V_1))$ is the induced subgraph of G by V_1. Let $E_1 \subset E$. Let $V_1 \subset V$ be all the vertices V which appear in the edges E_1. Then $G(E_1) := (V_1, E_1)$ is the induced subgraph of G by E_1.

A digraph $G = (V, E)$ is called *disconnected*, if the induced directed graph $H = (V, E_{\text{undir}})$ is disconnected. (*In the previous version* H was called *split*.) That is, there exists a decomposition of V to disjoint nonempty subsets $V = V_1 \cup V_2$ such that $G = G(V_1) \cup G(V_2)$. Otherwise G is called *connected*. For each digraph $G = (V, E)$ there exists a decomposition $V = \cup_{i=1}^{k} V_k$ to nonempty disjoint sets, called *connected components* of G, such that $G(V) = \cup_{i=1}^{k} G(V_i)$, where each $G(V_i)$, (a component of G), is connected. This decomposition is unique up to permutation of $V_1, \ldots, V_k$. *Note* that a digraph with one vertex only is always connected.

A digraph G is called *strongly connected* if for any two distinct vertices $u, v \in G$ there exists a path P from u to v, i.e. $V(P) = \{u_0 = u, u_1, \ldots, u_l = w\}$. Note that a graph having two vertices and exactly one edge from one vertex to another is connected but not strongly connected. If G is the maximal directed graph of an undirected graph H, then G is strongly connected if and only if H is connected.

Let W be a closed path on a digraph $G = (V, E)$, i.e. $V(W) = \{v_0, v_1, \ldots, v_l = v_0\}$. Denote by $G(W)$ the subgraph consisting of all vertices and all edges in W. Then it is straightforward to see that $G(W)$ is strongly connected.

A directed graph G is called *acyclic*, if G has no cycles. Note that G is acyclic iff it has no closed walks. (An *undirected* graph which is acyclic is called a *forest*.)

Proposition 2.14 *Let $G = (V, E)$ be an acyclic digraph. Then V decomposes to a disjoint union of k nonempty subsets $V_1, \ldots, V_k$ with the following properties: For each $v \in V_1$ $\deg_{inn}(v) = 0$, i.e. v does not have incoming edges. Assume that $k \geq 2$, i.e. G contains non-isolated vertices. Then for each $2 \leq i \leq k$ V_i consists of all u in V such that:*
(a) There exists $v(u) \in V_0$ and a path of length $i - 1$ connecting $v(u)$ to u.
(b) If there exists a path from $v \in V_0$ to u then its length is at most $i - 1$.

Proof. We first show that V_1 is nonempty. Take any vertex $v_0 \in V$. If $\deg_{\text{in}}(v_0) =$ then $v_0 \in V_1$. If $\deg_{\text{in}}(v_0) > 0$ there exists $v_{-1} \in V$ such that $(v_{-1}, v_0) \in E$. Suppose we can continue this process j steps, i.e. we have a walk W on G, such that $V(W) = \{v_{-j}, \ldots, v_{-1}, v_0\}$. Since G is acyclic W is a path. Suppose that $\#V = n$. Then $j \leq n - 1$. So this process must stop at some $j \leq n - 1$, i.e. $v_{-j} \in V_1$.

Define V_1 the set of all $v \in V$ such that $\deg_{\text{in}}(v) = 0$. Assume that $V_1 \neq V$. Let $v_0 \in V \backslash V_1 (= V - V_1)$, i.e. v_0 is any vertex in V which is not in V_1. From the above arguments there exists a path W on G, such that $V(W) = \{v_{-j}, \ldots, v_{-1}, v_0\}$ and $v_j \in V_1$. Let i be the maximal value of all possible values of j. Then v_0 belongs to V_{i+1}. Since V is finite the maximal possible value of i is k. □

Definition 2.15 *Let $G = (V, E)$ be acyclic digraph. A vertex $v \in V$ is called initial if $\deg_{\text{in}}(v) = 0$ and is called terminal if $\deg_{\text{out}}(v) = 0$. Any nonterminal vertex is called transient.*

Assume the conditions of Proposition 2.14. Then V_1 is the set of initial vertices. V_k is a subset of terminal vertices. Any V_i may contain a terminal vertex. A vertex $v \in V$ is initial and terminal iff v is isolated: $\deg_{\text{in}}(v) = \deg_{\text{out}}(v) = 0$.

Example: Let G be the following digraph on three vertices $\langle 3 \rangle$ and the edges $E = \{(1, 2), (1, 3), (2, 3)\}$. Then G is acyclic. $V_1 = \{1\}, V_2 = \{2\}, V_3 = \{3\}$.

The induced directed graph with the same vertices and undirected edges consists of one 3 cycle $(1, 2), (2, 3), (3, 1)$.

Reduced Graphs: We now discuss the structure of connected digraphs. Assume that $G = (V, E)$ is a connected digraph. On the set of vertices V of G we introduce the following relation $\sim$:

a. $v \sim v$ for any $v \in V$.

b. For $u, v \in V, u \neq v$ $u \sim v$ iff there exists a path (walk) in G connecting u to v and v to u.

It is straightforward to show that $\sim$ is an equivalence relation

$$u \sim u,\ u \sim v \iff v \sim u,\ u \sim v \ \texttt{and}\ v \sim w \Rightarrow u \sim w.$$

Hence V decomposes to a disjoin union of m nonempty equivalence classes $V = \cup_{i=1}^{m} V_i$. So for each V_i any two members $u, v \in V_i$ are connected in G, while for two disjoint V_i, V_j, and any $u \in V_i, v \in V_j$ either u is not connected to v or v is not connected to u, (or both). That is each $G(V_i)$ is a maximal strongly connected subgraph of G. Thus G is strongly connected iff $m = 1$. The *reduced* graph $G_{\rm rdc} = (V_{\rm rdc}, E_{\rm rdc})$ is given as follows. $V_{\rm rdc} = \{\{V_1\}, \ldots, \{V_m\}\}$. That is the vertices of $G_{\rm rdc}$ are the equivalence classes. $G_{\rm rdc}$ is loopless. For $i \neq j$ $(\{V_i\}, \{V_j\}) \in E_{\rm rdc}$ iff there exist $u \in V_i, v \in V_j$ such that u is connected to v. Since each V_i is an equivalence class it follows that $G_{\rm rdc}$ is acyclic. (Recall that a subgraph of a digraph induced by a cycle is strongly connected.) Since G is connected it follows that $G_{\rm rdc}$ is connected. Thus $G_{\rm rdc}$ gives the exact information on the "communication" between the maximal strongly connected subgraphs of G. Proposition 2.14 gives the structure of $G_{\rm rdc}$. The equivalence class V_i is called initial, terminal or transient, according to the status of the vertex $\{V_i\}$ in the acyclic graph $G_{\rm rdc}$.

Theorem 2.16 *Let $G = (V, E)$ be a strongly connected digraph. Let p be the* gcd *(the greatest common divisor) of the lengths of all cycles of G.*
(a) $p = 1$ if and only if there exits a positive integer N such that for any $u, v \in V$ and any $m \geq N$ there exists a walk W on G from u to v of length length m.
(b) $p \geq 2$ if an only if the following conditions satisfied. It is possible to divide V to p nonempty disjoint sets $V_1, \ldots, V_p$ such that the following conditions are satisfied. First,

$$E \cap (V_i \times V_{i+1}) \neq \emptyset \ \texttt{for}\ i = 1, \ldots, p, \quad E = \cup_{i=1}^{p} E \cap (V_i \times V_{i+1}). \tag{2.2}$$

Here $V_{p+1} \equiv V_1$. Second, there exists a positive integer N such that for each $m \geq N$ and any $u, v \in V_i$ there exists a a walk from u to v in pm steps for each $i = 1, \ldots, p$.

See for example [17, Chap 3] for a proof of this theorem using matrix tools.

We give a short proof of Theorem 2.16 using the following well known theorem.

Theorem 2.17 *Let $0 < a_1 < \ldots < a_k$ be k positive integers whose gcd is $p \in \mathbb{N}$. Then there exists $N \in \mathbb{N}$ such so that for any $m \geq N$ there exists $b_1, \ldots, b_k$ nonnegative integers such that $mp = b_1 a_1 + \ldots + b_k a_k$.*

Proof. By considering $a_i' = \frac{a_i}{p}, i = 1, \ldots, k$ it is enough to prove the theorem for $p = 1$. Since for $k = 1$ $p = a_1 = 1$ the theorem in this case is trivial since $m = ma_1$. Assume that $k > 1$. If $a_1 = 1$ then again the theorem is trivial since $m = ma_1 + 0a_2 + \cdots + 0a_k$. So assume assume $2 \leq a_1 < \ldots < a_k$.

Since the gcd of $a_1, \ldots, a_k$ is 1 it is known that there exists integers $c_1, \ldots, c_k$ such that

$$c_1 a_1 + c_2 a_2 + \cdots + c_k a_k = 1. \tag{2.3}$$

For a_1, a_2 one applies the Euclid algorithm to find (a_1, a_2) the gcd of (a_1, a_2). For example, for $a_2 = 3, a_2 = 17$ we have $6 \cdot 3 + (-1) \cdot 17 = 1$. Apply Euclid algorithm: $17 = 5 \cdot 3 + 2; 3 = 1 \cdot 2 + 1 \Rightarrow 17 = 5 \cdot 3 + (3 - 1) = 6 \cdot 3 - 1$. For $k > 2$ one first applies the Euclid algorithm to find the $g_2 = (a_1, a_2)$. Then apply Euclid algorithm to find $g_3 = (g_2, a_3)$ and cetera.

Since $2 \leq a_1 < \ldots < a_k$ and $c_1, \ldots, c_k$ are integer we must have at least one negative integer c_i and one positive integer c_j. Let $L := \max(-c_1, \ldots, -c_k)$ and $A := \sum_{i=1}^{k} a_i$. We

claim that any $m \geq LA^2$ is expressible as a nonnegative linear combination of $a_1, \ldots, a_k$ with nonnegative integer coefficients.

First note that if m is divisible by A, i.e. $m = tA$, then $m = ta_1 + ta_2 + \ldots ta_k$ and expressed m as a nonnegative linear combination of $a_1, \ldots, a_k$ with nonnegative integer coefficients. Thus it is left to consider the case where $m = tA + j$ for some $j = 1, \ldots, A-1$. The assumption that $m \geq LA^2$ yields that we can assume that $t \geq LA$. Multiply (2.3) by j to deduce that

$$m = tA + j = t\sum_{i=1}^{k} a_i + j\sum_{i=1}^{k} c_i a_i = \sum_{i=1}^{k}(t + jc_j)a_j.$$

Since $c_j \geq -L, j \leq A-1$ it follows that $(t + jc_j) \geq LA - (A-1)L = L > 0$. □

Proof of Theorem 2.16.

Fix a vertex $v \in V$. Consider the following closed walks in G starting and ending at v: $\mathcal{W} := \{W_1, \ldots, W_M\}$. First consider all cycles starting and ending at v. Now consider all closed walks which decomposed to two cycles in G. Continue in this manner until any cycle in G appears at least in one of this walks. It is not difficult to show that the gcd of all lengths of the walks in $\mathcal{W}$ is equal to the gcd of all cycles in G.

Assume first the case $p = 1$. Theorem 2.17 yields that there exists $N_1 \in \mathbb{N}$ such that for any $m \geq N_1$ there exists a closed walk W of length m starting and ending at v, which is obtained by using the set of walks in $\mathcal{W}$.

Since G is strongly connected there exists a path of length $P(v, w)$ from u to w, $(u \neq w)$, of length $n - 1 := \#V - 1$ at most. To get u to w we first go from u to v in a path $P(u, v)$, if $u \neq v$, then we take a closed walk of any length $\geq N_1$ around v and then we take a path $P(v, w)$. This shows that we can go from any vertex u to any vertex u' in a walk of length $m \geq N_1 + 2(n-1)$.

Assume now that the gcd of all cycles G is $p \geq 2$. Since any closed walk W in G can be decomposed to a sum of cycles it follows that any closed walk in G is divisible by p. Let $u \neq w$. Consider two walks W_1, W_2 from u to w. Complete each walk to a closed walk on from u to itself by taking a fixed path from w to u. Since each closed walk is divisible by p it follows that that the difference of lengths of W_1 and W_2 is divisible by p.

Fix a vertex v as above. The arguments of the proof of the theorem for $p = 1$ yield that there exists N_1 so that for any $m \geq N_1$ there is a closed walk from v to v in mp steps. For $i = 1, \ldots, p$ let V_i be all vertices in $u \in V$ such there exists a walk of from v to u of length l, where $l - i$ divisible by p. Note that $v \in V_p$. Next note that all vertices $u \in V$ such that $(w, u) \in E$, for some $w \in V_p$ form exactly the set of vertices V_1. Since G is strongly connected $\deg_{out}(v) > 0$ hence $V_1 \neq \emptyset$. Now V_2 is the set of vertices $u \in V$ such that $(w, u) \in E$ for some $w \in V_2$. Since G is strongly connected and V_1 is nonempty it follows that V_2 is not empty. Continue in the same manner to deduce that $V_1, \ldots, V_{p-1}$ are not empty. Clearly V_i is only connected to V_{i+1} for $i = 1, \ldots, p$. The rest of the theorem follows easily. □

We discuss briefly some aspects of Theorem 2.17. It is enough to consider the case $2 \leq a_1 < \ldots < a_k$ such that $a_1, \ldots, a_k$ are *coprime*, i.e. the gcd of $a_1, \ldots, a_k$ is 1. The problem of expressing $m \in \mathbb{N}$ as a linear combination of $a_1, \ldots, a_k$ with nonnegative integers is called the *coin problem*. Can one express the quantity of money m using only k types of coins of denomination $a_1, \ldots, a_k$? The largest number $f(a_1, \ldots, a_k)$ that *can not* be expressed as a linear combination of $a_1, \ldots, a_k$ with nonnegative integers is called is called the *Frobenius* number. For $k = 2$ Sylvester (1884) showed that $f(a_1, a_2) = (a_1 - 1)(a_2 - 1) - 1$. Explicit solutions for $k = 3$ are known: Selmer and Beyer 1978, Rödseth 1978, Greenberg 1988. No closed-form solution is known for $k > 3$. For a big k the problem of finding the Frobenius number is hard (**NP**-hard). The following theorem of I. Schur gives more precise version of Theorem 2.17. We outline a short proof using elementary results in theory of one

complex variable.

Theorem 2.18 *Let $2 \leq a_1 < \ldots < a_k$ be $k \geq 2$ positive integers whose gcd is 1. For any $m \in \mathbb{N}$ let $\alpha_m \in \mathbb{Z}_+$ be the number of ways that m can be expressed as $m = b_1 a_1 + \ldots b_k a_k$ with $b_1, \ldots, b_k$ nonnegative integers. Then*

$$\lim_{m \to \infty} \frac{\alpha_m (k-1)! a_1 \ldots a_k}{m^{k-1}} = 1. \tag{2.4}$$

Proof. For $q \in \mathbb{N}$ consider the polynomial $s_q(x) := 1 - x^q$. The roots of this polynomial are all $q-th$ roots of unity:

$$x_{j,q} := e^{\frac{2\pi j \sqrt{-1}}{q}}, j = 0, \ldots, q-1.$$

Here $\sqrt{-1}$ stands for the *imaginary* complex number, i.e. $\sqrt{-1}^2 = -1$. Note that $x_{0,q} = 1$. Let $p(x) = \prod_{i=1}^k s_{a_k}(x)$. Then the roots of $p(x)$, are union of all $a_i - th$ roots of unity for $i = 1, \ldots, k$. Note that 1 is a root of p of multiplicity k. Since $a_1, \ldots, a_k$ are co-prime any other root ζ of $p(x)$ has multiplicity less than k. Let $\zeta_0 = 1, \ldots, \zeta_l$ be all the distinct roots of $p(x)$. Let $m_j \in \mathbb{N}$ be the multiplicity of the root ζ_j for $j = 0, \ldots, l$. So $k = m_0 > m_j$ for $j = 1, \ldots, l$. Thus $p(x) = (-1)^k \prod_{j=0}^l (x - \zeta_j)^{m_j}$. Consider the rational function $r(x) := \frac{1}{p(x)}$. Note that $r(0) = 1$ and $r(x)$ is analytic in the unit disk $|x| < 1$. Hence $r(x)$ has power series expansion around $x = 0$ (Maclaurin expansion), with $\alpha_i := \frac{r^{(i)}(0)}{i!}$ for $i = 0, \ldots$. Use the geometric expansion $\frac{1}{1-t} = \sum_{i=0}^\infty t^i$ to deduce the identity

$$r(x) = \sum_{i=0}^\infty \alpha_i x^i = \prod_{j=1}^k \sum_{n=0}^\infty x^{n a_j}.$$

Hence each α_m is a nonnegative integer, and for $m \in \mathbb{N}$ α_m is the number of ways that m can be represented as a nonnegative linear combination $b_1 a_1 + \ldots + b_k a_k$ with nonnegative integers $b_1, \ldots, b_k$. Hence m is expressible as such a linear combination iff $\alpha_m > 0$. It is left to prove that the limit in (2.4) exists and equal to 1. The partial fraction decomposition of $r(x)$ is of the form

$$r(x) = \sum_{j=0}^l \sum_{i=1}^{m_j} \frac{A_{ji}}{(1 - x\zeta_j^{-1})^i}. \tag{2.5}$$

Note that

$$A_{0k} = \lim_{x \to 1} \frac{(1-x)^k}{p(x)} = \prod_{i=1}^k \lim_{x \to 1} \frac{(1-x)}{1 - x^{a_i}} = \frac{1}{a_1 \ldots a_k} > 0 \quad \texttt{(L'Hopital rule)}.$$

It is left to show that the contribution to the Maclarin coefficients of $r(x)$ by the terms given in (2.5) is dominated the Maclarin coefficients of $\frac{A_{0k}}{(1-x)^k}$ for high enough power of x. Recall the Newton binomial for $s \in \mathbb{N}$:

$$(1-t)^{-s} = \sum_{i=0}^\infty \binom{i+s-1}{s-1} t^i.$$

Thus the contribution of $\frac{A_{0k}}{(1-x)^k}$ to the $m - th$ Maclarin coefficient is of order $\frac{A_{0k} m^{k-1}}{(k-1)!}$. Use the fact that $A_{0k} > 0, k > m_j, |\zeta_j| = 1$ for each $j > 0$ to induce that the contributions from other terms in (2.5) to the Maclarin coefficients of $r(x)$ is of order m^{k-2} at most. Hence (2.4) holds. □

Let $G = (V, E)$ be an undirected connected graph. Let $G_{\max} = (V, E_{\max})$ be the directed graph induced by converting each undirected edge uv to two directed edges $(u, v), (v, u)$ in the opposite directions. Then $G_{\max}$ has a cycle of length 2: $\{u, v, u\}$. Use Theorems 2.16 and 2.10 to deduce:

Corollary 2.19 *Let $G = (V, E)$ be a connected undirected graph. Then exactly one of the following conditions hold:*
(a) There exists a positive integer N such that for any $u, v \in V$ and any $m \geq N$ there exists a walk W on G from u to v of length length m.
(b) G is bipartite, i.e. V is a union of two disjoint nonempty sets V_1, V_2 such that $E \subset E_1 \times E_2$. Let $u, v \in V_i$ for some $i \in \{1, 2\}$. Then any walk from u to v has an even length. Furthermore, there exists a positive integer N such that for any $m \geq N$ there exists a walk W on G from u to v of length length $2m$.

3 Random Graphs

3.1 Introduction

A random graph $\mathcal{G}_{n,p}$ consists of $2 \leq n \in \mathbb{N}$, a number $p \in [0, 1]$ and $\binom{n}{2}$ independent, identically distributed, Bernoulli random variables $X_{12}, \ldots, X_{1n}, X_{23}, \ldots, X_{(n-1)n}$, such that $\Pr(X_{ij} = 1) = p, \Pr(X_{ij} = 0) = 1 - p$ for $1 \leq i < j \leq n$. Let $\mathbf{X}_n := (X_{12}, \ldots, X_{(n-1)n}) \in \{0, 1\}^{\binom{n}{2}}$. Be a random vector. Then $G(\mathbf{X}_n) = (\langle n \rangle, E(\mathbf{X}_n))$ be the following undirected graph. For $1 \leq i < j \leq n$ the sedge (i, j) is in $E(\mathbf{X}_n)$ iff $X_{ij} = 1$. Then $\mathcal{G}_{n,p}$ is a sample space considering of all undirected graph $G = (\langle n \rangle, E) = G(\mathbf{X}_n) =$ on n vertices. The probability measure on $\mathcal{G}_{n,p}$ is given by $\Pr_{n,p}(G) = p^{\#E}(1-p)^{\binom{n}{2}-\#E}$. Let $\mathcal{A}$ be a property of an undirected graph. For example $\mathcal{A}$ is the property that the graph is *connected*. Thus $\mathcal{A}_n \subset \mathcal{G}_{n,p}$ is the subset of all connected undirected graphs on n vertices. $\mathcal{A}$ is called a *monotone* property (in p) if for the function $f_n(p) := \Pr_{n,p}(\mathcal{A}_n)$ is a monotone function. Intuitively, it is clear that connectivity property is an increasing function of p. For a general property $\mathcal{A}$ of undirected graphs, we let $\mathcal{A}_n := \{G \in \mathcal{G}_{n,p} : \ G \text{ has the property } \mathcal{A}\}$. As in [22] we let $\Pr_{n,p}(\mathcal{A}_n) = \Pr(G \in \mathcal{G}_{n,p} \text{ has } \mathcal{A})$. A function, (sequence), $p(n) > 0, n = 1, \ldots$, such that $p(n) \in [0, 1]$ for $n > N$, is called a *threshold function* for the property $\mathcal{A}$ if for any sequence $r(n) \in [0, 1], n \in \mathbb{N}$, the following conditions hold:
(a) $\lim_{n\to\infty} \frac{r(n)}{p(n)} = 0$ implies that $\lim_{n\to\infty} \Pr(G \in \mathcal{G}_{n,r(n)} \text{ has } \mathcal{A}) = 0$.
(b) $\lim_{n\to\infty} \frac{r(n)}{p(n)} = \infty$ implies that $\lim_{n\to\infty} \Pr(G \in \mathcal{G}_{n,r(n)} \text{ has } \mathcal{A}) = 1$.

The theory of random graphs was introduced by Paul Erdös and Alfred Renyi, see their seminal paper [6]. (There were other previous works on some aspects of random graphs.) Erdös and Renyi observed that for many natural properties $\mathcal{A}$ of undirected graphs there is a simple threshold function:

`Property`	`Threshold`
`Contains path of length` k	$p(n) = n^{-\frac{k+1}{k}}$,
`Is not planar`	$p(n) = \frac{1}{n}$,
`Contains a Hamiltonian path`	$p(n) = \frac{\log n}{n}$,
`Is connected`	$p(n) = \frac{\log n}{n}$,
`Contains a clique on` k `points`	$p(n) = n^{-\frac{2}{k-1}}$.

In the following two subsections we discuss the k-clique property and the property of an isolated vertex. (The property of non-isolated vertex is equivalent to connectivity.) In this two cases one can obtain more precise results then the threshold function, using the Poisson distribution.

3.2 k-clique property

Let $S \subset \langle n \rangle$ be a subset of k-elements. We denote this fact by $\#S = k$ and $S = \{i_1, i_2, \ldots, i_k\}, 1 \leq i_1 < i_2 < \ldots < i_k \leq n$. Let $G = (\langle n \rangle, E)$ be an undirected graph on n vertices. The G has an S-clique if $(i, j) \in E$ for any two distinct vertices $i, j \in S$. G has a k-clique if there exists $S \subset \langle n \rangle, \#S = 4$ such that G has an S-clique. Let $X_S : \mathcal{G}_{n,p} \to \{0, 1\}$

be the following Bernoulli random variable. For $G \in \mathcal{G}_{n,p}$ $X_S(G) = 1$ iff G contains a clique on the set S. We claim that $\Pr_{n,p}(X_S = 1) = p^{\binom{\#S}{2}}$. Indeed, to have a clique on S we must have all edges (i,j) for each pair $i,j \in S$. That is each random variable X_{ij} is equal to 1. Fix $k \geq 2$ and let $n \geq k$. Let $X_n = \sum_{S \subset \langle n \rangle, \#S=k} X_S$. Then $X_n : \mathcal{G}_{n,p} \to \mathbb{Z}_+$ is a random variable, such that $X_n(G)$ counts the number of k-cliques in a graph G on n vertices. Since the number of all possible k-cliques in a graph on n vertices is $\binom{n}{k}$, i.e. the number of all distinct choices of the sets of k elements in $\langle n \rangle$ it follows that

$$\mathrm{E}_p(X_n) = \sum_{S \subset \langle n \rangle, \#S=k} \mathrm{E}_p(X_S) = \binom{n}{k} p^{\binom{k}{2}} =$$

$$\frac{n(n-1)\dots(n-k+1)}{1 \cdot 2 \cdot \dots \cdot k} p^{\frac{k(k-1)}{2}} = (np^{\frac{k-1}{2}})^k \frac{1}{k!} \prod_{j=0}^{k-1} (1 - \frac{j}{n}). \tag{3.1}$$

Here we emphasized the fact that our expectation on $\mathcal{G}_{n,p}$ depends on p. Hence for sequences $r(n) \in [0,1] \in \mathbb{N}$ the following implication holds

$$\lim_{n \to \infty} n^{\frac{2}{k-1}} r(n) = a \in [0, \infty] \Rightarrow \lim_{n \to \infty} \mathrm{E}_{r(n)}(X_n) = \frac{a^{\binom{k}{2}}}{k!}. \tag{3.2}$$

In particular, if $\lim_{n \to \infty} n^{\frac{2}{k-1}} r(n) = 0, \infty$ then expected value of number of k-cliques tends to zero or infinity respectively. Thus the threshold function $p(n) = n^{-\frac{2}{k-1}}$ essentially gives the following rough information. If $\lim_{n \to \infty} n^{\frac{2}{k-1}} r(n) = 0, \infty$ then the probability of having no k-clique is approaching to one or zero respectively as $n \to \infty$.

We first show that property (a) of the threshold function $p(n) = n^{-\frac{2}{k-1}}$ follows simply for the fact that the expected number of cliques tends to 0. For that we need the following lemma.

Lemma 3.1 *Let Ω be a countable sample space with a probability measure. Let $X : \Omega \in \{0\} \cup [1, \infty)$ be a random variable. Then* $\Pr(X \geq 1) \leq \mathrm{E}(X)$.

Proof.

$$\Pr(X \geq 1) = \sum_{\omega \in \Omega, X(\omega) \geq 1} \Pr(\omega) \leq \sum_{\omega \in \Omega, X(\omega) \geq 1} X(\omega) \Pr(\omega) =$$

$$0 \cdot \Pr(X = 0) + \sum_{\omega \in \Omega, X(\omega) \geq 1} X(\omega) \Pr(\omega) = \mathrm{E}(X).$$

□

Since $X_n \in \mathbb{Z}_+$ we deduce

$$\lim_{n \to \infty} n^{\frac{2}{k-1}} r(n) = 0 \Rightarrow 1 \geq \Pr_{n,r(n)}(X_0) = 1 - \Pr_{n,r(n)}(X_n \geq 1) \geq 1 - E_{n,r(n)}(X_n) \to 1,$$

and property (a) follows.

To show property (b) we use Chebyshev's inequality. Clearly $X_n = 0$ satisfies the inequality $|X_n - \mathrm{E}_{n,p}| \geq |\mathrm{E}_{n,p}(X_n)| = \mathrm{E}_{n,p}(X_n)$ for any $n \geq k$ and $p \in [0,1]$. Chebyshev's inequality, Theorem 1.1, yields $\Pr(|X_n - \mathrm{E}_{n,p}| \geq |\mathrm{E}_{n,p}(X_n)|) \leq \frac{\mathrm{Var}_{n,p}(X_n)}{\mathrm{E}_{n,p}(X_n)^2}$. Thus we need to show that

$$\lim_{n \to \infty} n^{\frac{2}{k-1}} r(n) = \infty \Rightarrow \lim_{n \to \infty} \frac{\mathrm{Var}_{n,r(n)}(X_n)}{\mathrm{E}_{n,r(n)}(X_n)^2} = 0. \tag{3.3}$$

Recall that $X_n = \sum_{S \subset \langle n \rangle, \#S=k} X_S$, where each X_S is Bernoulli. If for any $S \neq T$ X_S, X_T are independent then (1.8) yields that

$$0 \leq \mathrm{Var}_{n,p}(X_n) = \sum_{S \subset \langle n \rangle, \#S=k} \mathrm{Var}_{n,p}(X_S) \leq \sum_{S \subset \langle n \rangle, \#S=k} \mathrm{E}_{n,p}(X_S) = \mathrm{E}_{n,p}(X_n),$$

and (3.3) will easily follows. However X_S, X_T are independent if $\#(S \cap T) \leq 1$. Hence we need to use second equality of (1.6), rather than (1.8). Note that since S and T are sets of a fixed size k and $n \to \infty$ it follows that for most of pairs $S, T \subset \langle n \rangle$ of cardinality k $S \cap T = \emptyset$. Hence most of the pairs X_S, X_T are independent Bernoulli variables, for large n. That is in (1.6) $\mathrm{Cov}(X_S, X_T) = 0$ for $\#(S \cap T) \leq 1$. Therefore (3.3) holds.

We now prove (3.3) for $k = 4$. Since the k-clique property is increasing in p. (The larger probability, it is more likely you have more edges in the graph, and hence bigger chances to have a k-clique.) It is enough to show property (b) under the condition

$$r(n) \leq \frac{\log n}{n^{\frac{2}{3}}} \quad \texttt{for } n \in \mathbb{N}, \quad \texttt{and } \lim n^{\frac{2}{3}} r(n) = \infty. \tag{3.4}$$

First observe that X_S, X_T are independent if $S \cap T = \emptyset$ since S and T are disjoint set. Assume next that $\#(S \cap T) = 1$. Then the set of edges on a complete graph on S and T are disjoint sets. Hence X_S, X_T are independent. In both cases $\mathrm{Cov}(X_S, X_T) = 0$.

Assume next that $\#(S \cap T) \in \{2, 3, 4\}$. Since X_S, X_T are Bernoulli, $\mathrm{E}(X_S), \mathrm{E}(X_T) \geq 0$ and $X_S X_T$ is Bernoulli. Hence

$$\mathrm{Cov}_{n,p}(X_S, X_T) = \mathrm{E}_{n,p}(X_S X_T) - E_{n,p}(X_S) E_{n,p}(X_T)) \leq$$
$$\mathrm{E}_{n,p}(X_S X_T) = \Pr{}_{n,p}(X_S X_T = 1) = \Pr{}_{n,p}(X_S = 1, X_T = 1).$$

The case $\#(S \cap T) = 2$. Then $X_S = X_T = 1$ means that we have 12 edges in the complete graph on S and T, of which 1 edge is a joint edge. Thus we have 11 edges. Hence $\Pr{}_{n,p}(X_S = 1, X_T = 1) = p^{11}$. The number of such S, T is found as follows. First choose the 2 joint vertices in $S \cap T$. Then number of such vertices in $\binom{n}{2}$. Then choose the remaining 2 vertices in S. The number of such choices is $\binom{n-2}{2}$. Then choose the other 2 vertices in T. Their number is $\binom{n-4}{2}$. Thus the total number of 4-sets S, T satisfying the condition $\#(S \cap T) = 2$ is $\binom{n}{2}\binom{n-2}{2}\binom{n-4}{2} \leq n^6$. Hence

$$\sum_{S,T \subset \langle n \rangle, \#S = \#T = 4, \#(S \cap T) = 2} \mathrm{Cov}_{n,r(n)}(X_S, X_T) \leq n^6 (\frac{\log n}{n^{\frac{2}{3}}})^{11} = \frac{(\log n)^{11}}{n^{\frac{4}{3}}} \to 0.$$

The case $\#(S \cap T) = 3$. Then the complete graphs on $S \cap T$ have three common edges. So the number of total edges in the complete graphs on S and T is $12 - 3 = 9$. Hence $\Pr{}_{n,p}(X_S = 1, X_T = 1) = p^9$. The number of such S, T is found as follows. First choose the 3 joint vertices in $S \cap T$. Then number of such vertices in $\binom{n}{3}$. Then choose the remaining vertex in S. The number of such choices is $n - 3$. Then choose the remaining vertex in T. Their number is $n - 4$. Thus the total number of 4-sets S, T satisfying the condition $\#(S \cap T) = 3$ is $\binom{n}{3}(n-3)(n-4) \leq n^5$. Hence

$$\sum_{S,T \subset \langle n \rangle, \#S = \#T = 4, \#(S \cap T) = 3} \mathrm{Cov}_{n,r(n)}(X_S, X_T) \leq n^5 (\frac{\log n}{n^{\frac{2}{3}}})^9 = \frac{(\log n)^9}{n} \to 0.$$

Thus

$$0 \leq \frac{\mathrm{Var}_{n,r(n)}(X_n)}{\mathrm{E}_{n,r(n)}(X_n)^2} = \frac{1}{\mathrm{E}_{n,r(n)}(X_n)^2} \Big(\sum_{S \subset \langle n \rangle, \#S = 4} \mathrm{Var}_{n,r(n)}(X_S) +$$
$$\sum_{S,T \subset \langle n \rangle, \#S = \#T = 4, \#(S \cap T) \in \{2,3\}} \mathrm{Cov}_{n,r(n)}(X_S, X_T) \Big) \leq \frac{\mathrm{E}_{n,r(n)}(X_n) + \frac{(\log n)^{11}}{n^{\frac{4}{3}}} + \frac{(\log n)^9}{n}}{\mathrm{E}_{n,r(n)}(X_n)^2} \to 0.$$

This shows (3.3) for $k = 4$, and completes the proof that $p(n) = n^{-\frac{2}{k-1}}$ is a threshold function for k-clique for the case $k = 4$.

Actually a stronger result holds.

Theorem 3.2 *Assume the equality in the first part of (3.2) with $a \in (0,\infty)$. Then X_n converges in probability to Poisson distribution* $\mathrm{Pu}(b)$ *with $b = \frac{a\binom{k}{2}}{k!}$. That is the probability that a random graph will have exactly j k-cliques is $e^{-b}\frac{b^j}{j!}$ for any $j \in \mathbb{Z}_+$.*

So if $a \to 0$ we obtain that with probability 1 a random graphs has no k-clique. If $a \to \infty$ then with probability 1 the random graph contains j k-cliques for any $j \geq 1$.

To prove the above theorem one need to use Theorem 1.6. That is, we need to show that

$$\lim_{n\to\infty} \mathrm{E}_{n,r(n)}\Big(\sum_{S_1,\ldots,S_j\subset\langle n\rangle,\#S_1=\ldots=\#S_j=k,S_i\neq S_l \text{ for } i\neq l} X_{S_1}\ldots X_{S_j}\Big) = b^j \quad \texttt{for each } j = 2,\ldots$$

Note that if $S_i \cap S_l = \emptyset$ for $i \neq l$ then $\mathrm{E}(X_{S_1}\ldots X_{S_j}) = \mathrm{E}(X_{S_1})\ldots\mathrm{E}(X_{S_j})$. Hence one has to show that the contribution of other terms is negligible. To show that one does similar computations as we did for the case $k = 2$ for $\mathrm{Var}(X_n)$ which is equivalent to consider the case $j = 2$.

3.3 Isolated vertices and connectivity

Theorem 3.3 *Let $c \in \mathbb{R}$ and $p(n) = \frac{\log n + c}{n}, n \in \mathbb{N}$. Then $\lim_{n\to\infty} \Pr(G \in \mathcal{G}_{n,p(n)}$ does not have an isolated vertex $) = e^{-e^{-c}}$.*

Proof. Let X_i be the Bernoulli variable corresponding to the event A_i : the vertex i is isolated. Then $\Pr_{n,p}(X_i = 1) = (1-p)^{n-1}$. Note that

$$(1-p(n))^{n-1} = e^{(n-1)\log(1-p(n))} = e^{-p(n)(n-1)+O((n-1)p(n)^2)} =$$
$$e^{-\log n - c + O(\frac{(\log n)^2}{n})} = \frac{e^{-c}}{n}\big(1 + O(\frac{(\log n)^2}{n})\big).$$

Let $X_n = \sum_{i=1} X_i$ be the random variable on $\mathcal{G}_{n,p}$, such that $X_n(G)$ is the number of isolated vertices in G. Then $\mathrm{E}_{n,p(n)}(X_n) = \big(1 + O(\frac{(\log n)^2}{n})\big)e^{-c} \to e^{-c}$. We claim that X_n converge in probability to the Poisson distribution $\mathrm{Pu}(b)$ with $b = e^{-c}$. We use Theorem 1.6. Note $\Pr_{n,p}(X_{i_1} = 1,\ldots,X_{i_k} = 1) = (1-p)^{(n-1)k-\binom{k}{2}}$, when $1 \leq i_1 < \ldots < i_k \leq n$. Indeed, this is the event that the given k vertices are isolated. Consider the complete graph K_n on n vertices. The degree of each vertex is $n-1$. So there are exactly $(n-1)k$ edges that coming out of these k vertices. $\binom{k}{2}$ edges are common to both vertices in these groups. Hence the total number of (distinct) edges that are connected to these k vertices is $(n-1)k - \binom{k}{2}$. Hence $\mathrm{E}_{n,p(n)}(X_{i_1} = 1,\ldots,X_{i_k} = 1) \approx \big(\frac{b}{n}\big)^k$. There are exactly $\binom{n}{k}$ of choices of k distinct vertices out of n. Hence

$$\lim_{n\to\infty} \mathrm{E}_{n,p(n)}\Big(\sum_{1\leq i_1<\ldots<i_k\leq n} X_{i_1}\ldots X_{i_k}\Big) = \frac{b^k}{k!}, \quad \texttt{for } k = 1,2,\ldots$$

Note that the event $G \in \mathcal{G}_{n,p(n)}$ does not have an isolated vertex is equivalent to $X_n = 0$. As $\Pr(\mathrm{Pu}(b) = 0) = e^{-b}$ we deduce the theorem. □

It can be shown that under the assumptions of Theorem 3.3 when $n \to \infty$ the the probability that G does not have an isolated vertex is equal to the probability that G is connected [6]. That is, under the assumptions of Theorem 3.3 $\lim_{n\to\infty} \Pr(G \in \mathcal{G}_{n,p(n)}$ is connected $) = e^{-e^{-c}}$. Therefore $p(n) = \frac{\log n}{n}$ is the threshold function for the connectivity property.

4 Matrices and Graphs 2-7-05

Let S be a set. In these notes S can be the set $\{0,1\}$, or the set of natural integers $\mathbb{N} := \{1,2,\ldots\}$, the set of integers $\mathbb{Z} = \{0,\pm 1,\pm 2,\ldots\}$, nonnegative integers $\mathbb{Z}_+ := \{0,1,2,\ldots\}$, real numbers $\mathbb{R}$ and complex numbers $\mathbb{C}$.

Denote by $S^{m\times n}$ the set of $m\times n$ matrices $A = (a_{ij})_{i,j=1}^{m,n}$, where the entries $a_{ij} \in S$:

$$A = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & ... & a_{mn} \end{pmatrix}$$

The transpose of A, denoted by $A^\top$ is an $n\times m$ matrix with $A^\top = (a_{ji})_{j,i=1}^{n,m}$. Note that $(A^\top)^\top = A$. Let $A = (a_{ij}), B = (b_{ij}) \in \mathbb{R}^{m\times n}$. Then $B \geq A \iff b_{ij} \geq a_{ij}$ for $i = 1,\ldots,m, j = 1,\ldots,n$.

A is called a square matrix (of order n) if A has the equal number of rows and columns (equal to n). A square matrix is called symmetric if $A^\top = A$.

Assume that A is a square matrix with complex entries $A = (a_{ij}) \in \mathbb{C}^{n\times n}$ The *trace* of A, denoted by $\operatorname{tr} A$, is the sum of all diagonal elements of : $\operatorname{tr} A := \sum_{i=1}^n a_{ii}$. For any $A \in \mathbb{C}^{n\times n}$ and a positive integer k, $A^k = A\cdot\ldots\cdot A$, where the product is taken k-times. It is agreed that $A^0 := I_n$, the $n\times n$ *identity* matrix. ($I = (\delta_{ij}) \in \{0,1\}^{n\times n}$, where $\delta_{ij} = 1$ iff $i = j$.) Note that if A is symmetric, i.e. $A = A^\top$, then A^k is also symmetric.

It is convenient to store graphs, as well to study certain graph properties, by using the matrices and their properties. First consider undirected graph $G = (V,E)$. Assume that $\#V = n$. Label the vertices V as $\{1,\ldots,n\}$, i.e. identify V with $\langle n\rangle$. Associate with G the matrix $A(G) = (a_{ij}) \in \{0,1\}^{n\times n}$ as follows. $a_{ij} = 1$ if $(i,j) \in E$. Otherwise $a_{ij} = 0$. Since G does not have loops $a_{ii} = 0, i = 1,\ldots,n$. Note that $A(G)$ is symmetric, i.e. $a_{ij} = a_{ji}$ for all $i,j = 1,\ldots,n$. ($A(G)^\top = A(G)$.) Vice versa, any symmetric $A = (a_{ij}) \in \{0,1\}^{n\times n}$ with zero diagonal induces a unique graph $G = G(A)$ on the set of vertices $\langle n\rangle$. Namely $ij \in E \iff a_{ij} = 1$.

Next consider digraph $G = (V,E)$. Assume that $\#V = n$. Label the vertices V as $\{1,\ldots,n\}$, i.e. identify V with $\langle n\rangle$. Then with G we associate the matrix $A(G) = (a_{ij}) \in \{0,1\}^{n\times n}$ as follows. $a_{ij} = 1$ if $(i,j) \in E$. Otherwise $a_{ij} = 0$. Vice versa, any $A = (a_{ij}) \in \{0,1\}^{n\times n}$ a unique digraph $G = G(A)$ on the set of vertices $\langle n\rangle$. Namely $(i,j) \in E \iff a_{ij} = 1$.

Note that if $A \in \{0,1\}^{n\times n}$ is a symmetric $0-1$ matrix with zero diagonal, we can view $G(A)$ either as an undirected graph, or the maximal directed graph corresponding to some undirected graph.

We now show how to *read* some graph properties from the matrices. Let $A = (a_{ij})_{i,j=1}^{m,n} \in \mathbb{C}^{m\times n}$. Then r_i and c_i are called the $i-th$ row and the $i-th$ column sum of A respectively:

$$r_i := \sum_{j=1}^n a_{ij}, \quad c_j := \sum_{i=1}^m a_{ij}, \quad j = 1,\ldots,n, i = 1,\ldots,m.$$

Assume that A is a square matrix. If A is symmetric than $r_i = c_i$. If $A = A^\top \in \{0,1\}^{n\times n}$ with zero diagonal, i.e. A represents an undirected graph G on n vertices, then $r_i = \deg(i)$ is the degree of the vertex i. If $A \in \{0,1\}^{n\times n}$, i.e. A represents a digraph G on n vertices, then $r_i = \deg_{\text{out}}(i), c_i = \deg_{\text{in}}(i)$.

Lemma 4.1 *Let G be a directed or undirected graph on n vertices. Let $A = (a_{ij}) \in \{0,1\}^{n\times n}$ be the representation matrix of G. For $k \in \mathbb{N}$ let $A^k := (a_{ij}^{(k)})_{i,j=1}^n$. Then $A^k \in \mathbb{Z}_+^{n\times n}$. $a_{ij}^{(k)}$ is the number of walks on G from the vertex i to the vertex j of length k. In particular, $\operatorname{tr} A^k$ is the number of closed walks on G of length k.*

Proof. Recall that $A^k = AA^{k-1}$ for each $k \in \mathbb{N}$. From the definition of matrix multiplication it follows

$$a_{ij}^{(k)} = \sum_{l=1}^{n} a_{il} a_{lj}^{(k-1)}, \quad i, j = 1, \ldots, n. \tag{4.1}$$

We first show by induction that the entries of any $A^k, k \in \mathbb{N}$ are nonnegative integers. Since $A \in \{0,1\}^{n \times n}$ each entry of A is either 0 or 1, hence in $\mathbb{Z}_+$. Assume by induction that the result hold for $k = p$. Let $k = p+1$ and use the formula (4.1) for this k. Since the products and the sum of nonnegative integers is a nonnegative integer it follows that every entry of A^{p+1} is a nonnegative integer.

We now show that the number of walks of from i to j of length k in G is given by by $a_{ij}^{(k)}$. Assume first that G is digraph. For $k = 1$ this is equivalent to the definition of the incidence matrix $A = A(G)$ of the digraph G. Assume by induction that the result hold for $k = p$. Let $k = p+1$. Let $W = \{i_0 = i, i_1, \ldots, i_{p+1} = j\}$ be walk on G of length $p+1$. Note that such walk exists iff $a_{ii_1} a_{i_1 i_2} \ldots a_{i_p j} = 1 \iff a_{ii_1} a_{i_1 i_2} \ldots a_{i_p j} \neq 0$. Denote by $\mathcal{W}_{p+1}(i,j) \supset \mathcal{W}_{p+1}(i,l,j)$ the set of all walks on G from i to j in $p+1$ steps and the subset of such walks so the first step in this walk is from i to l. Note that either if this sets can be an empty set. ($\mathcal{W}_{p+1}(i,j) = \emptyset$ iff there is no walk in G of length $p+1$ from i to j.) Then $\mathcal{W}_{p+1}(i,j) = \cup_{l=1}^{n} \mathcal{W}_{p+1}(i,l,j)$. $\mathcal{W}_{p+1}(i,l,j) \neq \emptyset \iff a_{il} = 1$ and $\mathcal{W}_p(l,j) \neq \emptyset$. By the induction assumption $\#\mathcal{W}_p(l,j) = a_{lj}^{(p)}$. Hence $\#\mathcal{W}_{p+1}(i,l,j) = a_{il} a_{lj}^{(p)}$. Thus

$$\#\mathcal{W}_{p+1}(i,j) = \sum_{l=1}^{n} \#\mathcal{W}_{p+1}(i,l,j) = \sum_{l=1}^{n} a_{il} a_{lj}^{(p)} = a_{ij}^{(p+1)}.$$

Therefore $\#\mathcal{W}_k(i,j) = a_{ij}^{(k)}$ as claimed. In particular $\#\mathcal{W}_k(i,i) = a_i^{(k)}$ is the number of walks starting and ending at i. (Any walk that ends at the starting point is called a *closed* walk, or *periodic* walk, (of period k). Hence $\operatorname{tr} A^k$ is the number of periodic walks on G with period k.

For an undirected graph the proof is the same, since the walk on $G = (\langle n \rangle, E)$ corresponds to the walk on the oriented graph where $a_{ij} = a_{ji} = 1$ iff $ij \in E$. □

The above theorem is standard, can be found in [15, I.3]

Corollary 4.2 *Let G be directed or undirected graph on n vertices. Then G has a cycle of odd length iff* $\operatorname{tr} A^{2k-1}$ *is at least* 1 *for some* $k = 1, \ldots, \lfloor \frac{n+1}{2} \rfloor$.

Proof. Assume first that G is a digraph. Assume first that there exists an odd cycle of length $2k-1$ ($\leq n$). Clearly, $k \leq \lfloor \frac{n+1}{2} \rfloor$. Hence $\operatorname{tr} A^{2k-1}$ is at least 1. Suppose now that $\operatorname{tr} A^{2k-1}$ is at least 1 for some $k \leq \lfloor \frac{n+1}{2} \rfloor$. Hence there exists a closed walk W on G of length $2k-1$. It is straightforward to show that any closed walk W on G can be decomposed as a "sum" of cycles. The length of the walk W is the sum of the lengths of the cycles. Since the length of the closed walk A is odd, there must be at least one odd cycle in any decomposition of W to a "sum" of cycles.

Assume that $G = (\langle n \rangle, E)$ is an undirected graph. Then a *semi-cycle* on G is defined as a closed walk of length 2 : $W = (iji)$ where $ij \in E$. Then any closed walk on G decomposes to "sum" of cycles and semi-cycles and the proof in this case follows as for the digraph. □

Remark 4.3 *Recall that undirected or directed graph $G = (\langle n \rangle, E)$ has no odd cycles iff it is bipartite. It is not difficult to find a fast polynomial (quadratic) algorithm in n, which finds if G is bipartite or not.*

However to find if a given undirected or directed graph $G = (\langle n \rangle, E)$ has an even cycle is a hard (NP-complete) problem.

Lemma 4.4 *Let G be a digraph on n vertices and let $A \in \{0,1\}^{n\times n}$ be its representation matrix. Then G is strongly connected iff all the entries of $B = (b_{ij})_{i,j=1}n := A^0 + A^1 + \ldots + A^{n-1}$ are positive.*

Proof. Recall that $A^k \in \mathbb{Z}_+^{n\times n}$ for each $k \in \mathbb{Z}_+$. Hence $B \geq A^k$ for any $k \in [0, n-1]$. In particular $B \geq A^0 = I_n$, hence all diagonal entries of B are positive. Assume first that G is strongly connected. Then for any two distinct vertices $i \neq j \in \langle n \rangle$ there exists a path P of length k which connects i to j. Since all the vertices in P are distinct $k \leq n$. So $b_{ij} \geq a_{ij}^{(k)} \geq 1$.

Assume now that all the entries of B are positive. Let $i \neq j$ be two distinct vertices. From the definition of B it follows that there exists $k \in [1, n-1]$ such $a_{ij}^{(k)} > 0$. Since $a_{ij}^{(k)}$ is a nonnegative integer it follows that $a_{ij}^{(k)} \geq 1$. Thus we have a walk of length k from i to j. Hence we have a path of length k at most from i to j. Thus G is strongly connected. □

$A \in \mathbb{C}^{m\times n}$ is called a *block* matrix if if $A = (A_{ij})_{i,j=1}^{pq}$ and each $A_{ij} \in \mathbb{C}^{m_i\times n_j}$ for $i = 1, \ldots, p,\ j = 1, \ldots q$:

$$A = \begin{pmatrix} A_{11} & A_{12} & ... & A_{1q} \\ A_{21} & A_{22} & ... & A_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ A_{p1} & A_{p2} & ... & A_{pq} \end{pmatrix}, \tag{4.2}$$

$$A_{ij} \in \mathbb{C}^{m_i\times n_j}, i = 1, \ldots, p, j = 1, \ldots, q, \sum_{i=1}^{p} = m_i, \sum_{j=1}^{q} n_j = n.$$

A square matrix $A \in \mathbb{C}^{n\times n}$ is called *block diagonal* is if it is a block matrix where each diagonal block is a square matrix and all off-diagonal elements are zero matrices, i.e. their entries are all equal to zero:

$$A = \mathrm{diag}(A_1, \ldots, A_q) = \oplus_{j=1}^{q} A_j := \begin{pmatrix} A_1 & 0 & ... & 0 \\ 0 & A_2 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & A_q \end{pmatrix},$$

$$A_i \in \mathbb{C}^{n_i\times n_i}, i = 1, \ldots, q, \sum_{j=1}^{q} n_j = n.$$

Note that if a digraph (undirected graph) is a disjoint union of q graphs: $G = \cup_{j=1}^{q} G_j$ then $A(G) = \oplus_{j=1}^{q} A(G_j)$.

Assume that a digraph $G = (V, E)$ is acyclic. Then Proposition 2.14 yields that the representation matrix $A(G)$ is the following $k \times k$ block matrix:

$$A = \begin{pmatrix} 0 & A_{12} & A_{13} & A_{14} & ... & A_{1k} \\ 0 & 0 & A_{23} & A_{24} & ... & A_{2k} \\ \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & ... & A_{(k-1)k} \\ 0 & 0 & 0 & 0 & ... & 0 \end{pmatrix},\ A_{ij} \in \mathbb{C}^{n_i\times n_j}, i = 1, \ldots, k, j = i+1, \ldots, k. \tag{4.3}$$

Moreover $A_{i(i+1)} \neq 0$ for $i = 1, \ldots, k$.

Assume that the digraph $G = (V, E)$ is strongly connected and is periodic, i.e. not aperiodic. That is, the conditions of Theorem 2.16 with $p > 1$. In the decomposition $V = \cup_{i=1}^{p} V_i$ assume that $\#V_i = n_i, i = 1, \ldots, p$. Then the representation matrix $A(G)$ is

the following $p \times p$ block matrix:

$$A = \begin{pmatrix} 0 & A_{12} & 0 & 0 & ... & 0 \\ 0 & 0 & A_{23} & 0 & ... & 0 \\ \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & ... & A_{(p-1)p} \\ A_{p1} & 0 & 0 & 0 & ... & 0 \end{pmatrix}, \ A_{i(i+1)} \in \mathbb{C}^{n_i \times n_{i+1}}, i = 1, \ldots, p, (p+1 \equiv 1). \tag{4.4}$$

Any directed (undirected) G is bipartite if it has a representation matrix of the above form for $p = 2$.

Reduced Graphs and reducible matrices: Let $G = (V, E)$ be a digraph. Assume that G is not strongly connected. Recall the definition of the reduced graph given in §2.2. Let $V_1, \ldots, V_m$ be the decomposition of V to nonempty union such that each $G(V_i)$ is a maximal connected component. Then $G_{\mathrm{rdc}} = (V_{\mathrm{rdc}}, E)$ is acyclic. We can always assume that we can rename the vertices of the maximal strongly connected components of G such that we do not have edges from V_i to V_j if $j < i$. In particular the vertex $\{V_1\} \in V$ represents an initial vertex in G_{rdc} while $\{V_k\}$ represents the terminal state in G_{rdc}. It is possible that G_{rdc} have other terminal (initial) states $\{V_i\}$. Then the representation matrix $A(G)$ has the following upper diagonal block form induced by the nonempty sets V_i of cardinality n_i for $i = 1, \ldots, k$:

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & ... & A_{1(k-1)} & A_{1k} \\ 0 & A_{22} & A_{23} & ... & A_{2(k-1)} & A_{2k} \\ \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & ... & A_{(k-1)(k-1)} & A_{(k-1)k} \\ 0 & 0 & 0 & ... & 0 & A_{kk} \end{pmatrix}, \tag{4.5}$$

$$A_{ij} \in \mathbb{C}^{n_i \times n_j}, i = 1, \ldots, k, j = i+1, \ldots, k.$$

Note that if $n_i = 1$, i.e. V_i consists of one vertex, then either $A_{ii} = (0)$ or $A_{ii} = (1)$. If $n_i > 1$ then $G(V_i)$ is a strongly connected graph having more than one vertex.

Vice versa, a matrix $A \in \mathbb{C}^{n \times n}$ is called *reducible* if by *renumbering* the rows and the columns (in the same way) A has the form (4.5). Equivalently, there exists a permutation matrix $X \in \{0, 1\}^{n \times n}$ such that $XAX^\top$ is of the form (4.5). Recall that X is a permutation matrix if each row and column of X has exactly one entry equal to 1 and all other entries are equal to 0. Equivalently, X it is the representation matrix of a graph on n vertices which is a disjoint union cycles. Recall that $X^\top X = XX^\top = I_n$, i.e. $X^{-1} = X^\top$.

$A \in \mathbb{C}^{n \times n}$ is called a diagonal matrix if it is a square matrix, whose all off-diagonal entries are 0:

$$\mathrm{diag}(d_1, d_2, ..., d_n) = \begin{pmatrix} d_1 & 0 & \ldots & 0 & 0 \\ 0 & d_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & d_n \end{pmatrix}$$

Example: $\mathrm{diag}(3, -2, 7) = \begin{pmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 7 \end{pmatrix}$. Note that the identity matrix is a diagonal matrix with all diagonal entries equal to 1. More general, an $m \times n$ matrix $A = (a_{ij})_{i,j=1}^{m,n}$ is called a diagonal matrix and denoted by $\mathrm{diag}(d_1, \ldots, d_{\min(m,n)})$ if $a_{ij} = 0$ for $i \neq j$ and $a_{ii} = d_i$ for $i = 1, \ldots, \min(m, n)$.

Assume that $A = \mathrm{diag}(A_1, \ldots, A_q)$ is a block diagonal matrix. Then it is straightforward to show that $A^k = \mathrm{diag}(A_1^k, \ldots, A_q^k)$ for any $k \in \mathbb{N}$. Clearly $A^0 = \mathrm{diag}(A_1^0, \ldots, A_q^0)$. It is easy to raise to the power $k \in \mathbb{N}$ a diagonal matrix $D = \mathrm{diag}(d_1, \ldots, d_n)$: $D^k = \mathrm{diag}(d_1^k, \ldots, d_n^k)$. This basic fact can be used for computing the powers of A for "most"of matrices square

matrices, since most of the quare matrices are similar to a diagonal matrix. That is for most $A \in \mathbb{C}^{n\times n}$ there exists an invertible matrix $X \in \mathbb{C}^{n\times n}$ a a diagonal matrix $\Lambda := \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \in \mathbb{C}^{n\times n}$ such that $A = X\Lambda X^{-1}$. (Note that $XX^{-1} = X^{-1}X = I_n$.) Hence $A^k = X\Lambda^k X^{-1}$ for any $k \in \mathbb{N}$.

Given any matrix $A = (a_{ij})_{i,j}^n \in \mathbb{C}^{n\times n}$ we associated with A a digraph $G(A) := (\langle n\rangle, E)$ as follows: $(i,j) \in E \iff a_{ij} \neq 0$. We call G the *graph* associated with A.

5 Markov chains on digraphs 2-16-05

5.1 Basic properties 2-16-05

A nonnegative matrix $P = (p_{ij})_{i,j=1}^n \in [0,\infty)^{n\times n}$ is called *stochastic*, or some times *row* stochastic, if $\sum_{j=1}^n p_{ij} = 1$ for $i = 1, \ldots, n$. That is each row of a stochastic matrix is a probability vector. Equivalently, if $\mathbf{e} := (1, \ldots, 1) \in \mathbb{R}^n$ is the vector whose all coordinates is equal to 1, then $P \geq 0$ and $P\mathbf{e} = \mathbf{e}$.

Let $G = (\langle n\rangle, E)$ a digraph. As in [10] is convenient to view the vertices $\langle n\rangle$ of G as vertices $V := \{v_1, \ldots, v_n\}$ or states $S := \{s_1, \ldots, s_n\}$. Assume that G has the following property: For any $i \in \langle n\rangle$ $\deg_{\mathrm{out}}(i) \geq 1$.

Imagine a particle jumps from vertex to vertex at discrete times measured by nonnegative integers $m = 0, 1, \ldots,$. For each $m \in \mathbb{Z}_+$ let X_m be the random variable that gives the position of a particle at time X_m. So $\Pr(X_m = i) = \mu_i^{(m)}, i = 1, \ldots, n$. Here $X_m = i$ means that the particle is at time m is at the vertex v_i (s_i). Hence $\mu^{(m)} := (\mu_1^{(m)}, \ldots, \mu_n^{(m)})$ is a *row* probability vector. The sequence of random variables $X_0, X_1, \ldots,$ is called a *process*, or random process. Denote by $\Pr(X_m|X_0, X_1, \ldots, X_{m-1})$ the probability of the distribution of X_m knowing the values of $X_0, \ldots, X_{m-1}$. More specifically, $\Pr(X_m = i_m|X_0 = i_0, X_1 = i_1, \ldots, X_{m-1} = i_{m-1})$ denotes the probability of a particle being in location $i_m \in \langle n\rangle$, provided the the particle was at the place i_l at time $l = 0, \ldots, m-1$.

The process $X_0, X_1, \ldots,$ is called *Markov* process, or Markov *chain* if the following property holds. For each $m \geq 1$ the conditional probability $\Pr(X_m|X_0, X_1, \ldots, X_{m-1})$ is equal to the conditional probability $\Pr(X_m|X_{m-1})$. More precisely

$$\Pr(X_m = i_m|X_0 = i_0, X_1 = i_1, \ldots, X_{m-1} = i_{m-1}) = \Pr(X_m = i_m|X_{m-1} = i_{m-1})$$

for any $i_0, \ldots, i_m \in \langle n\rangle$. The exact values of the $\Pr(X = i_m|X_{m-1} = i_{m-1})$ are descrived as follows. Let $P_m := (p_{ij,m})_{i,j=1}^n, m = 1, \ldots$ be a sequence $n \times n$ stochastic matrices such that $G(P_m) := (\langle n\rangle, E_m)$ is a subgraph of G for each $m \in \mathbb{N}$. That is $E_m \subset E, m \in \mathbb{N}$. Then

$$\Pr(X = i_m|X_{m-1} = i_{m-1}) = p_{i_{m-1}i_m,m}, \quad i_{m-1}, i_m \in \langle n\rangle,\ m = 1, \ldots. \tag{5.1}$$

That is if we know that at time $m-1$ the particle is at the place i_{m-1} then at time m the particle will only be at the places i_m where $(i_{m-1}, i_m) \in E$. This follows from the condition that $G(P_m) \subset G$. The stochasticity of P_m equivalent to the fact that the particle at time m jumps to some vertex in G from the vertex he was at time $m-1$. The general form (5.1) is called *inhomogeneous* Markov chain. See [10, §2,p'13-14] for a simple example for inhomogeneous Markov chain model describing the weather of Gothenburg in summer and winter.

We claim that

$$\mu^{(m)} = \mu^{(m-1)}P_m, \quad m = 1, \ldots. \tag{5.2}$$

Indeed

$$\Pr(X_m = j) = \sum_{i=1}^n Pr(X_{m-1} = i)\Pr(X_m = j|X_{m-1} = i) = \sum_{i=1}^n \mu_i^{(m-1)}p_{ij,m}.$$

In particular we have $\mu^{(m)} = \mu^{(0)}P_1P_2\ldots P_m$, for any $m \in \mathbb{N}$.

Markov chain is called *homogeneous* if $P_m = P$ for all $m \in \mathbb{N}$ and some fixed stochastic P. In that case P is called the *transition* matrix of the Markov chain. Usually one assumes that $G = G(P)$. Then

$$\mu^{(m)} = \mu^{(0)} P^m, \quad m = 1, \dots . \tag{5.3}$$

See See [10, §2] for examples of simple homogeneous Markov chains.

5.2 Computer simulation of homogeneous Markov chains 2-18-05

Let $P = (p_{ij})_{i,j}^n \in [0,1]^{n \times n}$ be a stochastic matrix. How do we simulate the homogeneous Markov chain generated by P on the graph $G = G(P)$? To do that one needs to assume that we have a program that generates a random variable U with uniform distribution on $[0,1]$. That is U takes only values in the unit interval $[0,1]$, i.e. $0 \le U \le 1$ and for any $t \in [0,1]$ $\Pr(U \le t) = t$.

Assume that the random variable X_m has value $i \in \langle n \rangle$ at time $m \in \mathbb{Z}_+$. Equivalently, at time m our particle is at the state s_i. For $j = 0, 1, \dots, n$ define $q_{ij} \in [0,1]$ as follows

$$q_{i0} = 0, \quad q_{ij} = \sum_{l=1}^{j} p_{il}, \ \texttt{for } j = 1, \dots, n, \quad \texttt{for } i = 1, \dots, n. \tag{5.4}$$

Apply the subroutine which generates the random variable U. Then there exists exactly one $j \in \langle n \rangle$ such that $q_{i(j-1)} \le U < q_{ij}$. (*Show it!*) Now let $X_{m+1} = j$. That is the particle jumped from the state s_i at time m to the state s_j at time $m+1$. See for more details [10, §3].

5.3 Stationary distributions

A row probability vector $\mu = (\mu_1, \dots, \mu_n)$ is called a *stationary distribution* for a stochastic matrix $P = (p_{ij})_{i,j}^n$ if $\mu P = \mu$. Assume that we have a homogeneous Markov chain on $G = G(P)$ induced by P. Let $X_0, X_1, \dots$ be the Markov process $X_0, X_1, \dots$ of random walks on G induced by P. Suppose that the initial distribution of X_0 is given by the stationary distribution μ: $\Pr(X_0 = i) = \mu_i, i = 1, \dots, n$. Then (5.3) implies that $\mu^{(m)} = \mu$ for any $m \in \mathbb{N}$. That is $X_0, X_1, \dots$ are identically distributed. We then have the following *natural* problems in theory of Markov chains:
1. *Do stationary distributions always exist*?
2. Under what conditions there exists a *unique* stationary distribution?
3. Suppose that P admits a unique stationary distribution μ. Under what conditions the Markov process $X_0, X_1, \dots$ converges to a unique random variable X with the stationary distribution μ?

The answer Problem 1 is yes. We know the answers to Problems 2 and 3. To state the answers we need the following definitions.

Let $G = G(P)$. Then P is called *irreducible* if G is strongly connected. An irreducible P is called *aperiodic* or *periodic* if G is a strongly connected graph, which is aperiodic or periodic respectively. P is called *reducible* if G is not strongly connected. Let G_{rdc} be the reduced graph of G. Let $G(V_i)$ be a maximal strongly connected component of G. Then V_i is called a terminal subset if $\{V_i\}$ is a terminal vertex in G_{rdc}.

Proposition 5.1 *Let $P = (p_{ij})_{i,j=1}^n$ be a stochastic matrix. Let $G = G(P) = (\langle n \rangle, E)$ and assume that that $V \subset \langle n \rangle$ be a nonempty subset of $\langle n \rangle$. Denote by $P(V) := (p_{ij})_{i,j \in V}$ the square submatrix of P induced by the rows and columns of P which are in V. Then $P(V)$ is a substochastic matrix. That is, each entry of $P(V)$ is nonnegative and the sum of each row is at most 1. $P(V)$ is an irreducible stochastic matrix if and only if exactly on of the following conditions holds.*
(a) P is irreducible, i.e. G is strongly connected. Then $V = \langle n \rangle$.
(b) P is reducible, $G(V)$ is a maximal strongly connected component of G and $\{V\}$ is a terminal vertex in the acyclic reduced graph G_{rdc} corresponding to G.

Proof. Let $i \in V$. Then

$$\sum_{j \in V} p_{ij} \leq \sum_{j=1}^{n} p_{ij} = 1. \tag{5.5}$$

Hence $P(V)$ is substochastic. Note that $\sum_{j \in V} p_{ij} = 1$ iff $(i,j) \notin E$ for any $j \notin V$.

Suppose first that G is strongly connected. Let $V \subsetneqq \langle n \rangle$ be a nonempty subset. (Hence $n > 1$.) Since G is strongly connected there exist $i \in V$ and $k \notin V$ such that $(i,k) \in E$. Hence $p_{ik} > 0$. In particular (5.5) yields that $\sum_{j \in V} p_{ij} < 1$. Thus $P(V)$ is not stochastic. Thus we are left with the case $V = \langle n \rangle$. Then $P(V) = P$ is irreducible.

Assume that P is reducible, i.e. G is not strongly connected. Suppose first that $P(V)$ is an irreducible stochastic matrix. In particular $V \neq \langle n \rangle$. Then $G(V) = G(P(V))$ is strongly connected. Since $P(V)$ is stochastic it follows that for any $i \in V$ one has the equality sign in (5.5). Thus $i \in V, k \notin V \Rightarrow (i,k) \notin V$. That is there are no edges from V to $\langle n \rangle \backslash V$. Therefore $G(V)$ is maximal connected component of G and V is terminal.

Suppose now that $G(V)$ is a maximal strongly connected component of G. As $G(V) = G(P(V))$ $P(V)$ is irreducible. Suppose that V is terminal. So there are no edges from V to $\langle n \rangle \backslash V$. Hence for any $i \in V$ the equality sign in (5.5). Thus $P(V)$ is stochastic. □

Let $\alpha = (\alpha_1, \ldots, \alpha_n)$ be a probability vector (distribution) on $\langle n \rangle$. Then $\operatorname{supp} \alpha = \{i \in \langle n \rangle : \ \alpha_i > 0\}$ is the support of the distribution α. Denote by Π_n the set of all distributions $\alpha = (\alpha_1, \ldots, \alpha_n)$ on $\langle n \rangle$. For any $p \in \mathbb{N}$ let $\mathbf{x}_1, \ldots, \mathbf{x}_p \in \mathbb{R}^d$ be p vectors. Let $(\beta_1, \ldots, \beta_p) \in \Pi_p$ be a distribution on $\langle p \rangle$. Then $\mathbf{x} := \sum_{l=1}^{p} \beta_l \mathbf{x}_l \in \mathbb{R}^d$ is a convex combination of $\mathbf{x}_1, \ldots, \mathbf{x}_p$. Let

$$\operatorname{conv}(\mathbf{x}_1, \ldots, \mathbf{x}_p) := \{\mathbf{x} \in \mathbb{R}^d : \quad \mathbf{x} = \sum_{l=1}^{p} \beta_i \mathbf{x}_i \ \texttt{for all} \ (\beta_1, \ldots, \beta_p) \in \Pi_p\}.$$

Then $\operatorname{conv}(\mathbf{x}_1, \ldots, \mathbf{x}_p)$ is called the *convex hull* spanned by $\mathbf{x}_1, \ldots, \mathbf{x}_p$. It is a convex set. ($C \subset \mathbb{R}^d$ is called convex set if any convex combination of any two points is in C.)

Here is a probabilistic interpretation of $\operatorname{conv}(\mathbf{x}_1, \ldots, \mathbf{x}_p)$. Let $X : \Omega \to \mathbb{R}^d$ be a random variable such that $X(\Omega) = \{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$. Then $\mathrm{E}(X)$ is a point in $\mathbb{R}^d$. The set of all possible values of $\mathrm{E}(X)$ is given by $\operatorname{conv}(\mathbf{x}_1, \ldots, \mathbf{x}_p)$.

Similarly if $\alpha^1, \ldots, \alpha^p \in \Pi_n$ are p distributions on $\langle n \rangle$ then $\operatorname{conv}(\alpha^1, \ldots, \alpha^p)$ is the convex hull of row vectors spanned by $\alpha^1, \ldots, \alpha^p$. It is straightforward to show that each $\alpha \in \operatorname{conv}(\alpha^1, \ldots, \alpha^p)$ is a distribution on $\langle n \rangle$.

Theorem 5.2 *Let $P = (p_{ij})_{i,j}^n$ be a stochastic matrix. Let $G = G(P) = (\langle n \rangle, E)$ be the induced digraph by P. Then the following holds:*
1. Suppose that P is irreducible, i.e. G is strongly connected. Then P has a unique stationary distribution $\mu = (\mu_1, \ldots, \mu_n)$, and $\mu_i > 0$ for $i = 1, \ldots, n$.
(a) If G (or P) is aperiodic then for any distribution $\mu^{(0)}$ $\lim_{m \to \infty} \mu^{(0)} P^m = \mu$. That is the Markov process $X_0, X_1, \ldots$, converges to the unique random variable X given by the stationary distribution μ. ($\Pr(X = i) = \mu_i, i = 1, \ldots, n$.)
(b) G (or P) is periodic. Assume that $p \geq 2$ is the gcd of all cycles of G. Then for each $\mu^{(0)}$ each of the sequence $\mu^{(0)} P^{mp+i}, m = 1, 2, \ldots$ converges for $i = 0, 1, \ldots, p-1$. These limits depend of $\mu^{(0)}$ and i.
2. Assume that P is reducible, i.e. G is not strongly connected. Let G_{rdc} be the reduced graph of G with the vertices $\{V_1\}, \ldots, \{V_k\}$. Then for each terminal vertex $\{V_i\}$ in G_{rdc} there exists a unique distribution $\mu(V_i) \in \Pi_n$ with $\operatorname{supp} \mu(V_i) = V_i$. Assume that $\{V_{i_1}\}, \ldots, \{V_{i_t}\}$ are all the terminal vertices of G_{rdc}. Then the set of all stationary distributions of P is the convex hull spanned by $\mu(V_{i_1}), \ldots, \mu(V_{i_t})$. Hence P has a unique stationary distribution iff G_{rdc} has exactly one terminal vertex $\{V_{i_1}\}$.
(a) Assume that G_{rdc} has exactly one terminal vertex $\{V_{i_1}\}$ and suppose furthermore that $G(V_{i_1})$ is aperiodic. Then for any distribution $\mu^{(0)}$ $\lim_{m \to \infty} \mu^{(0)} P^m = \mu(V_{i_1})$. That is

the Markov process $X_0, X_1, \ldots$, converges to the unique random variable X given by the stationary distribution $\mu(V_{i_1})$.

(b) Assume that either $t = 1$ and $G(V_{i_1})$ is periodic or $t > 1$. Let $p_i \geq 1$ be the gcd of all cycles of $G(V_{i_j})$ for $j = 1, \ldots, t$. Let p be the smallest positive integer that is divisible by $p_1, \ldots, p_t$. Then for each $\mu^{(0)}$ each of the sequence $\mu^{(0)} P^{mp+i}, m = 1, 2, \ldots$ converges for $i = 0, 1, \ldots, p-1$. These limits depend of $\mu^{(0)}$ and i.

To prove this theorem we will need some results on the spectral properties of nonnegative matrices. The following corollary, which is deduced straightforward from the part (2b) of the is called the *mean ergodic theorem* for stochastic matrices.

Corollary 5.3 *Let $P \in [0,1]^{n \times n}$ be a stochastic matrix. Then $\frac{1}{m} \sum_{i=0}^{m-1} P^i$ converges to a stochastic matrix Q as $m \to \infty$. Q is a projection: $Q^2 = Q$, which satisfies $PQ = QP = Q$. Furthermore, for any distribution $\alpha \in \Pi_n$, $\lim_{m \to \infty} \frac{1}{m} \sum_{i=0}^{m-1} \alpha P^i = \alpha Q$ is a stationary distribution of P.*

6 Square matrices with nonnegative entries

6.1 Eigenvectors and eigenvalues of square complex matrices

(All the results of this subsection can be found in [15] and are part of Math 310 course in UIC.) Let $\mathbb{C}^{n \times n}$ be the algebra of $n \times n$ complex valued matrices. That is, we can define $A + B, AB, aA$ for any $A, B \in \mathbb{C}^{n \times n}, a \in \mathbb{C}$, which satisfy the *appropriate rules*. Then one can define the determinant function $\det : \mathbb{C}^{n \times n} \to \mathbb{C}$ which satisfies the following properties

$$\det aA = a^n \det A, \ \det AB = \det A \det B, \ \det I_n = 1, \texttt{for any } A, B \in \mathbb{C}^{n \times n}, a \in \mathbb{C}.$$

Moreover $\det : \mathbb{R}^{n \times n} \to \mathbb{R}$, i.e. the determinant of a real matrix is a real number, and $\det : \mathbb{Z}^{n \times n} \to \mathbb{Z}$, i.e. the determinant of a integer valued matrix is an integer. The great advantage of $\det A$ that it can be computed in a polynomial time, e.g. at most in $\frac{2n^3}{3}$ operations.

Assume that $A \in \mathbb{C}^{n \times n}$. $\det A \neq 0$ iff there exists $A^{-1} \in \mathbb{C}^{n \times n}$, called the *inverse* of A, such that $AA^{-1} = A^{-1}A = I_n$. $\det A = 0$ iff the rows (columns) of A are linearly independent. Equivalently, $\det A = 0$ iff the system $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution $\mathbf{x} \neq \mathbf{0}$. $\mathbf{0} \neq \mathbf{x} \in \mathbb{C}^n$ is called the right (left) eigenvector with the corresponding eigenvalue λ if $A\mathbf{x} = \lambda \mathbf{x}$ ($\mathbf{x}^\top A = \lambda \mathbf{x}^\top$). The eigenvalues of A are the complex roots of the *characteristic polynomial*

$$\det(zI_n - A) = z^n - (\operatorname{tr} A) z^{n-1} + \ldots + (-1)^n \det A = \prod_{i=1}^{n} (z - \lambda_i). \tag{6.1}$$

Hence

$$\operatorname{tr} A = \sum_{i=1}^{n} \lambda_i, \quad \det A = \lambda_1 \cdot \lambda_2 \cdots \lambda_n. \tag{6.2}$$

The set of all distinct eigenvalues of A, is denoted by $\operatorname{spec} A \subset \mathbb{C}$. Thus $\operatorname{spec} I_n = \{1\}$, since $\det(zI_n - I_n) = (z-1)^n$. Hence 1 is the unique root of the characteristic polynomial of multiplicity n. Thus I_n has n eigenvalues $\lambda_1 = \ldots = \lambda_n = 1$. It is customary to arrange the n eigenvalues of A in one the following two orders

$$\begin{gathered} \Re(\lambda_1) \geq \Re(\lambda_2) \geq \ldots \geq \Re(\lambda_n) \\ |\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|. \end{gathered} \tag{6.3}$$

The spectral radius of A, denoted by $\rho(A)$, is defined as the maximal modulus of all the eigenvalues of A. That is, if we arrange the eigenvalues of A with respect to the second of the above order, then $\rho(A) = |\lambda_1|$.

The problem with the eigenvalues of A that it may happen that some or all eigenvalues of a real matrix can be complex (not reals). For example $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ has two complex eigenvalues $\lambda_1 = \sqrt{-1}, \lambda_2 = -\sqrt{-1}$. The corresponding eigenvectors can not be chosen real.

It is convenient to define

$$\mathrm{Eig}(z, A) := \{\mathbf{x} \in \mathbb{C}^n : \quad A\mathbf{x} = z\mathbf{x}\}, \quad \texttt{for any } z \in \mathbb{C}.$$

Clearly $\mathrm{Eig}(z, A)$ is a subspace of A. Then $z \in \operatorname{spec} A \iff \dim \mathrm{Eig}(z, A) > 0$. If z is not an eigenvalue of A the $\mathrm{Eig}(z, A) = \{\mathbf{0}\}$ is the trivial subspace. For $\lambda \in \operatorname{spec} A$ $\mathrm{Eig}(\lambda, A)$ is the *eigenspace* of A corresponding to λ. Thus any $\mathbf{0} \neq \mathbf{x} \in \mathrm{Eig}(\lambda, A)$ is an eigenvector of A corresponding to λ. The dimension of the vector subspace $\mathrm{Eig}(\lambda, A)$, denoted by $\dim \mathrm{Eig}(\lambda, A)$, is called the *geometric multiplicity* of λ. The multiplicity $m(\lambda)$ of the root λ in the characteristic polynomial $\det(zI_n - A)$ is called the *algebraic multiplicity* of λ. It is known that $\dim \mathrm{Eig}(\lambda, A) \leq m(\lambda)$. λ is called *geometrically simple* if $\dim \mathrm{Eig}(\lambda, A) = m(\lambda)$. Otherwise λ is called a *defective* eigenvalue.

Assume that $\lambda_i \neq \lambda_j$ for $1 \leq i < j \leq m \leq n$ are m distinct eigenvalues of A. Let $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{C}^{n \times n}$ be the corresponding eigenvectors of A: $A\mathbf{x}_i = \lambda_i \mathbf{x}_i, ; i = 1, \ldots, m$. Then it is known that $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are linearly independent. This is equivalent to the statement that the dimension of the subspace $\mathbf{V}$, spanned by all eigenvectors vectors in of A, is equal to the sum of the dimensions of all different eigenspaces of A. In other notation $\mathbf{V} = \oplus_{\lambda \in \operatorname{spec} A} \mathrm{Eig}(\lambda, A)$.

Suppose that $m = n$, that is the characteristic polynomial of A has n distinct eigenvalues. (This is a generic situation. If one chooses at random all the n^2 entries of A then with probability 1 A will have n distinct eigenvalues.) This condition is equivalent to the assumption that each eigenvalue of A is algebraically simple. Then the square matrix $X := (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \in \mathbb{C}^{n \times n}$ whose $i-th$ column is $\mathbf{x}_i$ has nonzero determinant. That is X^{-1} exists. Then

$$A\mathbf{x}_i = \lambda_i \mathbf{x}_i, \ i = 1, \ldots, n \iff AX = X\Lambda \iff A = X\Lambda X^{-1}, \ \Lambda := \mathrm{diag}(\lambda_1, \ldots, \lambda_n). \tag{6.4}$$

Such an A is called *diagonable*. Recall that A is diagonable iff for any eigenvalue λ of multiplicity $m \geq 1$ in the characteristic polynomial of A, $\mathrm{rank}\,(\lambda I_n - A) = n - m$. If $A = X\Lambda X^{-1}$ then for any $k \in \mathbb{N}$ $A^k = X\Lambda^k X^{-1}$ and it is fairly simple to understand the behavior of $A^k, k = 1, \ldots$.

Let $G(A) = (V, E)$ be the digraph induced by A. Let $V = \cup_{i=1}^k V_i$ be a decomposition of V to a union of nonempty disjoint sets such that each $G(V_i)$ is a connected component of G. (That does not mean that $G(V_i)$ is strongly connected!) That is for $i \neq j$ there are no directed edges between V_i and V_j, and each undirected graph induced by $G(V_i)$ is connected. Then there exist a permutation matrix $Q \in \{0, 1\}^{n \times n}$ such that

$$QAQ^T = \mathrm{diag}(A_1, \ldots, A_k), \quad A_i \in \mathbb{C}^{n_i \times n_i}, \ i = 1, \ldots, k. \tag{6.5}$$

For simplicity of notation we assume that we renamed the vertices $\langle n \rangle$ such that $Q = I_n$, i.e. A is a block diagonal

$$A = \mathrm{diag}(A_1, \ldots, A_k) = \oplus_{i=1}^k A_i, \quad A_i \in \mathbb{C}^{n_i \times n_i}. \tag{6.6}$$

Then $\det(\lambda I_n - A) = \prod_{i=1}^k \det(\lambda I_{n_i} - A_i)$. To find the eigenvalues and the eigenvectors of A it is enough to find the eigenvalues and the eigenvectors of each A_i. More precisely we view $\mathbb{C}^n = \oplus_{i=1}^k \mathbb{C}^{n_i}$ as follows. Every $\mathbf{w} \in \mathbb{C}^n$ is viewed as $\mathbf{w}_1 \oplus \ldots \oplus \mathbf{w}_k$, where $\mathbf{w}_i \in \mathbb{C}^{n_i}$ for $i = 1, \ldots, k$. That is $\mathbf{w}^\top = (\mathbf{w}_1^\top, \mathbf{w}_2^\top, \ldots, \mathbf{w}_k^\top)$. (Remember that $\mathbf{w}$ and $\mathbf{w}_1, \ldots, w_k$ are

column vectors!) Then we can give the following description of the eigenspace $\mathrm{Eig}(\lambda, A)$:

$$\mathrm{Eig}(\lambda, \oplus_{i=1}^k A_i) = \oplus_{i=1}^k \mathrm{Eig}(\lambda, A_i), \quad \dim \mathrm{Eig}(\lambda, \oplus_{i=1}^k A_i) = \sum_{i=1}^k \dim \mathrm{Eig}(\lambda, A_i), \tag{6.7}$$

for any $\lambda \in \mathrm{spec}\, A$. (Actually the above formula hodls for any $\lambda \in \mathbb{C}$.) That is, any eigenvector of A corresponding to eigenvalue λ is of the form $\mathbf{y}^\top = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_k^\top)$ where the following conditions are satisfied. If $\lambda \notin \mathrm{spec}\ (A_i)$ then $\mathbf{y}_i = 0$. If $\lambda \in \mathrm{spec}\ (A_i)$ then either $\mathbf{y}_i$ is an eigenvector of A_i corresponding to λ or $\mathbf{y}_i = 0$. At least one of the vectors $\mathbf{y}_i$ must be nonzero. (For this $\mathbf{y}_i$ $\lambda \in \mathrm{spec}\, A_i$.)

Let $P \in \mathbb{R}_+^{n\times n}$, i.e. P is an $n \times n$ matrix with nonnegative matrices. Then P is stochastic iff $P\mathbf{1} = \mathbf{1}$, where $\mathbf{1} = (1, \ldots, 1)^\top$.

Theorem 6.1 *Let P be a stochastic matrix. Denote by $G(P) = (\langle n \rangle, E)$ the digraph induced by P. Then 1 is an eigenvalue of P, and each other eigenvalue λ of P satisfies $|\lambda| \leq 1$. Let λ be an eigenvalue of P of modulus 1, i.e. $|\lambda| = 1$. Then there exists a terminal vertex $\{V\}$ in the reduced graph G_{rdc} such that λ is an eigenvalue of the irreducible stochastic matrix $P(V)$ of period $p(V)$. Furthermore λ is a simple root of $\det(zI - P(V))$ and $\lambda^{p(V)} = 1$. Moreover any $p(V)$-root of 1 is an eigenvalue of $P(V)$ and hence of P. Furthermore eigenvalue of P of modulus 1 is geometrically simple.*

Proof. We show here that any eigenvalue λ of P satisfies $|\lambda| \leq 1$. Assume that $P\mathbf{x} = \lambda\mathbf{x}, \mathbf{x} \neq \mathbf{0}$. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top$ and assume that $|x_k| = \max_{i\in[1,n]} |x_i| > 0$. Then

$$|\lambda x_k| = |(P\mathbf{x})_k| = |\sum_{i=1}^n p_{ki}x_i| \leq \sum_{i=1}^n p_{ki}|x_i| \leq \sum_{i=1}^n p_{ki}|x_k| = |x_k|.$$

Divide by $|x_k|$ to obtain $|\lambda| \leq 1$. Other claims of the Theorem follow from Perron-Frobenius theorem and Theorem 5.2. We will prove the special case where P is symmetric in the next subsections. □

6.2 Spectral theory of real symmetric matrices

(Most of the results of this subsection can be found in [15] and are part of Math 310 course in UIC.) In this subsection we assume that $A \in \mathbb{R}^{n\times n}$ matrix is *symmetric*, i.e. if $A^\top = A$. Denote by $\mathrm{S}_n(\mathbb{R})$ the set of all $n \times n$ real symmetric matrices. Note that $\mathrm{S}_n(\mathbb{R})$ is a vectors space over $\mathbb{R}$. Then A has only real eigenvalues and is diagonable. In this case we assume that $\mathrm{Eig}(\lambda, A) \subset \mathbb{R}^n$, i.e. we consider only the real eigenvectors of A. Note that any right eigenvector of A is also left eigenvector of A corresponding to the same eigenvalue. Furhtermore, it is possible to choose the eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^n$ of A such that

$$A\mathbf{x}_i = \lambda_i\mathbf{x}_i,\ i = 1, \ldots, n,\ \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n,\ \mathbf{x}_i^\top\mathbf{x}_j = \delta_{ij},\ i, j = 1, \ldots, n. \tag{6.8}$$

Recall that δ_{ij} is the (i, j) entry of the identity matrix I_n. The set of n vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^n$ satisfying the conditions $\mathbf{x}_i^\top\mathbf{x}_j = \delta_{ij}$ for $i, j = 1, \ldots, n$ is called an *orthonormal basis* of $\mathbb{R}^n$. Let $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{n\times n}$ then the condition that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is an orthonormal set is equivalent to $X^\top X = I_n \iff X^{-1} = X^\top$, i.e. X is an *orthogonal matrix*. Denote by O_n the set, (group), of orthonormal matrices. Then (6.8) is equivalent to $A = X\Lambda X^\top$. Note that $X \in \mathbb{R}^{n\times n}$ is orthogonal matrix iff $(X\mathbf{x})^\top(X\mathbf{x}) = \mathbf{x}^\top\mathbf{x}$ for any $\mathbf{x} \in \mathbb{R}^{n\times n}$. Observe that $||\mathbf{x}|| := \sqrt{\mathbf{x}^\top\mathbf{x}}$ is a nonnegative number, which is positive unless $\mathbf{x} = 0$. $||\mathbf{x}||$ is called *norm* or *length* of $\mathbf{x}$. The maximal eigenvalue λ_1 has the maximal characterization

$$\lambda_1 = \max_{\mathbf{x}\neq 0} \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top\mathbf{x}} = \max_{||\mathbf{x}||=1} \mathbf{x}^\top A\mathbf{x}. \tag{6.9}$$

Equality is achieved iff $\mathbf{x}$ is an eigenvector corresponding to λ_1. Here is a quick argument. If $A = \Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ the proof is straightforward. In the general case let $\mathbf{y} = X^\top \mathbf{x}$ and note that $\mathbf{x}^\top A\mathbf{x} = \mathbf{y}^\top \Lambda \mathbf{y}$ and $\mathbf{x}^\top \mathbf{x} = \mathbf{y}^\top \mathbf{y}$ and the proof of this case follows from the previous case. Similarly

$$\lambda_n = \min_{\mathbf{x} \neq 0} \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top \mathbf{x}} = \min_{||\mathbf{x}||=1} \mathbf{x}^\top A\mathbf{x}.$$

Equality holds iff $\mathbf{x}$ is the eigenvector corresponding to λ_n. There is also an extremal characterization of λ_2:

$$\lambda_2 = \max_{\mathbf{x} \neq 0, \mathbf{x}_1^\top \mathbf{x} = 0} \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top \mathbf{x}} = \max_{||\mathbf{x}||=1, \mathbf{x}_1^\top \mathbf{x} = 0} \mathbf{x}^\top A\mathbf{x}. \tag{6.10}$$

Equality is achieved iff $\mathbf{x}$ is an eigenvector corresponding to λ_2 which is orthogonal to $\mathbf{x}_1$. The disadvantage of this characterization is that it requires the knowledge of $\mathbf{x}_1$. There are other characterizations which avoid it, but we will not give them here.

Let $G(A) = (V, E)$ be the graph induced by A. Let $V = \cup_{i=1}^k V_i$ be a decomposition of V to a union of nonempty disjoint sets such that each $G(V_i)$ is a connected component of G. Then there exist a permutation matrix $Q \in \{0,1\}^{n \times n}$ such that (6.5) holds. Furthermore each $A_i \in \mathrm{S}_{n_i}(\mathbb{R})$ is real symmetric. Assume as in the previous subsection that $Q = I_n$. The one has the decomposition (6.6) where $A_i \in \mathrm{S}_{n_i}(\mathbb{R}), i = 1, \dots, k$. To find each eigenspace $\operatorname{Eig}(\lambda, A)$ we use (6.7).

6.3 Singular value decomposition and l_p operator norms of matrices

Theorem 6.2 *Let $A \in \mathbb{R}^{m \times n}$. Then there exists orthogonal matrices $U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{n \times n}$ and a diagonal matrix $\Sigma := \operatorname{diag}(\sigma_1, \dots, \sigma_{\min(m,n)}) \in \mathbb{R}^{m \times n}$, with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(n,m)} \geq 0$, such that $A = U\Sigma V^\top$. Assume that $\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_{\min(m,n)} = 0$. Assume that $\sigma_i = 0$ for any $i > \min(m,n)$. Then $r = \operatorname{rank} A$. $\sigma_1^2 \geq \ldots \geq \sigma_r^2$ are all the positive eigenvalues of $AA^\top$ and $A^\top A$, arranged is a decreasing order. All other eigenvalues of $AA^\top$ and $A^\top A$ are equal to zero. Let $U = (\mathbf{u}_1, \dots, \mathbf{u}_m)$ and $V = (\mathbf{v}_1, \dots, \mathbf{v}_n)$. Then $\mathbf{u}_1, \dots, \mathbf{u}_m$ are the orthonormal eigenvectors of $AA^\top$ corresponding to the eigenvalues $\sigma_1^2, \dots, \sigma_m^2$. Furthermore $\mathbf{v}_i = \frac{1}{\sigma_i} A^\top \mathbf{u}_i$ for $i = 1, \dots, r$ is an orthonormal system in $\mathbb{R}^n$. $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ is any completion of $\mathbf{v}_1, \dots, \mathbf{v}_r$ to an orthonormal basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of $\mathbb{R}^n$. Similarly, let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be any orthonormal set of eigenvectors of $A^\top A$ corresponding to the eigenvalues $\sigma_1^2, \dots, \sigma_n^2$. Then $\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i$ for $i = 1, \dots, r$. $\mathbf{u}_{r+1}, \dots, \mathbf{u}_m$ is any completion of $\mathbf{u}_1, \dots, \mathbf{u}_r$ to an orthonormal basis $\mathbf{u}_1, \dots, \mathbf{u}_n$ of $\mathbb{R}^n$.*

Proof. Consider $B = AA^\top \in \mathbb{R}^{m \times m}$. Then B is symmetric and nonnegative definite, i.e. $\mathbf{x}^\top B\mathbf{x} = (A^\top \mathbf{x})^\top (A^\top \mathbf{x}) \geq 0$. Hence all the eigenvalues of B are nonnegative, and denote them by $\sigma_1^2 \geq \ldots \geq \sigma_m^2 \geq 0$. Note that $A^\top \mathbf{x} = 0 \Rightarrow B\mathbf{x} = 0$. Furthermore $B\mathbf{x} = 0 \Rightarrow \mathbf{x}^\top B\mathbf{x} = 0 = (A^\top \mathbf{x})^\top (A^\top \mathbf{x}) \Rightarrow A\mathbf{x} = 0$. Hence $\operatorname{rank} A = \operatorname{rank} B = r$. Therefore $\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_{\min(m,n)} = 0$. Let $B\mathbf{u}_i = \sigma_i^2 \mathbf{u}_i$ for $i = 1, \dots, m$, where $\mathbf{u}_1, \dots, \mathbf{u}_m$ is an orthonormal basis in $\mathbb{R}^m$. So $\sigma_i^2 = \mathbf{u}_i^\top B\mathbf{u}_i = (A^\top \mathbf{u}_i)^\top (A^\top \mathbf{u}_i)$. It now follows that $\mathbf{v}_i = \frac{1}{\sigma_i} A^\top \mathbf{u}_i$ for $i = 1, \dots, r$ is an orthonormal system in $\mathbb{R}^n$. Let $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ be any completion of $\mathbf{v}_1, \dots, \mathbf{v}_r$ to an orthonormal basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of $\mathbb{R}^n$. It is straightforward to show that $A = U\Sigma V^\top$. The other part of the theorem follows similarly. □

Corollary 6.3 *Let $A \in \mathrm{S}_n(\mathbb{R})$ and assume that $A = X \operatorname{diag}(\lambda_1, \dots, \lambda_n) X^\top, X \in \mathrm{O}_n$ be the spectral decomposition of A. The the sequence $\sigma_1 \geq \ldots \geq \sigma_n \geq 0$ of singular values of A is the rearranged sequence of $|\lambda_1|, \dots, |\lambda_n|$. Let $\sigma_i = |\lambda_j|$ then one can choose $\mathbf{v}_i = \mathbf{x}_j$ and $\mathbf{u}_i = \epsilon_i \mathbf{x}_j$, where $\epsilon_i = \frac{\lambda_j}{|\lambda_j|}$ if $\lambda_j \neq 0$ and $\epsilon_j = \pm 1$ if $\lambda_j = 0$.*

Proof. Use the identity $AA^\top = A^\top A = A^2 = X \operatorname{diag}(\lambda_1^2, \dots, \lambda_n^2) X^\top$. □

A function $\nu : \mathbb{R}^n \to \mathbb{R}_+$ is called a *norm* if the following properties hold
a. $\nu(\mathbf{x}) > 0$ for $\mathbf{x} \neq \mathbf{0}$ and $\nu(\mathbf{0}) = 0$. (*positivity*);
b. $\nu(a\mathbf{x}) = |a|\nu(\mathbf{x})$ for any $a \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$, (*homogeneity*);
c. $\nu(\mathbf{x}+\mathbf{y}) \leq \nu(\mathbf{x}) + \nu(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, (*triangle inequality*).

The most common examples of norms are the l_p norms

$$||\mathbf{x}||_p = ||\mathbf{x}^\top||_p := \big(\sum_{i=1}^n |x_i|^p\big)^{\frac{1}{p}}, \texttt{for } p \in [1,\infty), \quad ||\mathbf{x}||_\infty = ||\mathbf{x}^\top||_\infty := \max_{1\leq i\leq n} |x_i|. \tag{6.11}$$

(We remark that sometimes we use in the sequel l_p norms of row vectors.) The Euclidean norm $||\mathbf{x}||_2 = ||\mathbf{x}^\top||_2$ is equal to the norm $||\mathbf{x}|| = ||\mathbf{x}^\top|| = \sqrt{\mathbf{x}^\top\mathbf{x}}$ defined in §6.2.

View $A \in \mathbb{R}^{m\times n}$ as an operator $A : \mathbb{R}^n \to \mathbb{R}^m$ given by $\mathbf{x} \mapsto A\mathbf{x}$. The (operator) l_p norm of A is defined as

$$||A||_p := \max_{||\mathbf{x}||=1, \mathbf{x}\in\mathbb{R}^n} ||A\mathbf{x}||_p = \max_{\mathbf{0}\neq\mathbf{x}\in\mathbb{R}^n} \frac{||A\mathbf{x}||_p}{||\mathbf{x}||_p} \quad \texttt{for } A \in \mathbb{R}^{m\times n} \texttt{ and } p \in [1,\infty]. \tag{6.12}$$

$||\cdot||_p$ is a norm on $\mathbb{R}^{m\times n}$ such that $||I_n||_p = 1$ for any $n \in \mathbb{N}$. Note that from the definition of the operator norm it follows

$$\begin{aligned}
&||A\mathbf{x}||_p \leq ||A||_p||\mathbf{x}||_p \texttt{ for all } \mathbf{x} \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n},\\
&||AB||_p \leq ||A||_p||B||_p \texttt{ for all } A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times l},\\
&||\operatorname{diag}(d_1,\ldots,d_{\min(m,n)})||_p = \max_{i\in[1,\min(m,n)]} |d_i| \texttt{ for any } \operatorname{diag}(d_1,\ldots,d_{\min(m,n)}) \in \mathbb{R}^{m\times n},\\
&||A^{l+q}||_p \leq ||A^l||_p||A^q||_p \texttt{ for all } l,q \in \mathbb{Z}_+ \texttt{ and } A \in \mathbb{R}^{n\times n}.
\end{aligned} \tag{6.13}$$

Theorem 6.4 *For any $m \times n$ real valued matrix $A = (a_{ij})_{i,j=1}^{m,n}$ one has*

$$||A|| := ||A||_2 = \sigma_1(A),\ ||A||_1 = \max_{j\in[1,n]} \sum_{i=1}^m |a_{ij}|,\ ||A||_\infty = \max_{i\in[1,m]} \sum_{j=1}^n |a_{ij}|.$$

Proof. Clearly $||Q\mathbf{x}|| = ||\mathbf{x}||$ for any orthogonal matrix Q. Use singular value decomposition of A to deduce

$$\begin{aligned}
&||A|| = ||U\operatorname{diag}(\sigma_1,\ldots,\sigma_{\min(m,n)})V^\top|| =\\
&||\operatorname{diag}(\sigma_1,\ldots,\sigma_{\min(m,n)})V^\top|| = ||\operatorname{diag}(\sigma_1,\ldots,\sigma_{\min(m,n)})|| = \sigma_1.
\end{aligned}$$

For $\mathbf{x} = (x_1,\ldots,x_n)^\top \in \mathbb{R}^n$ we have

$$||A\mathbf{x}||_1 = \sum_{i=1}^m |\sum_{j=1}^n a_{ij}x_j| \leq \sum_{i,j=1}^{m,n} |a_{ij}|\,|x_j| = \sum_{j=1}^n (\sum_{i=1}^m |a_{ij}|)|x_j| \leq \big(\max_{j\in[1,n]} \sum_{i=1}^m |a_{ij}|\big)||\mathbf{x}||_1.$$

Hence $||A||_1 \leq \max_{j\in[1,n]} \sum_{i=1}^m |a_{ij}|$. Choose $\mathbf{x} = \mathbf{e}_k = (\delta_{1k},\ldots,\delta_{nk})^\top$, where $k = \arg\max_{j\in[1,n]} \sum_{i=1}^m |a_{ij}|$, i.e. an index for which the maximum is achieved. Then $||A\mathbf{e}_k||_1 = \max_{j\in[1,n]} \sum_{i=1}^m |a_{ij}|$ $||\mathbf{e}_k||_1 = \max_{j\in[1,n]} \sum_{i=1}^m |a_{ij}|$. Hence the second equality of the theorem follows. The third equality of the theorem follows similarly. □

Corollary 6.5 *Let $A \in \mathbb{R}^{m\times n}$. Then $||A^\top||_1 = ||A||_\infty$. Let $P \in [0,1]^{n\times n}$ be a stochastic matrix. Then $||P||_\infty = ||P^\top||_1 = 1$. In particular $||P\mathbf{x}||_\infty \leq ||\mathbf{x}||_\infty$ and $||\mathbf{y}^\top P||_1 \leq ||\mathbf{y}^\top||_1$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Equality hold for any $\mathbf{y} \in \mathbb{R}^n_+$.*

Definition 6.6 *A sequence of real numbers $a_i, i \in \mathbb{N}$ is called subadditive if $a_{i+j} \leq a_i + a_j$ for all $i, j \in \mathbb{N}$.*

The following lemma is fundamental in many areas of mathematics.

Lemma 6.7 *Let $a_i, i \in \mathbb{N}$ be a real subadditive sequence. Assume that $\frac{a_i}{i}, i \in \mathbb{N}$ is bounded from below. Then the sequence $\frac{a_i}{i}$ converges to a limit $a \in \mathbb{R}$ and $a \le \frac{a_j}{j}$ for each $j \in \mathbb{N}$.*

Proof. Assume that $a_0 = 0$. Then the subadditivity conditions extends to $a_i, i \in \mathbb{Z}_+$. Note that $a_{ij} = a_{j+(i-1)j} \le a_j + a_{(i-1)j} \le \ldots \le ia_j$. Hence $\frac{a_{ij}}{ij} \le \frac{a_j}{j}$. Let $a = \liminf_{i\to\infty} \frac{a_i}{i}$. Since $\frac{a_i}{i}, i \in \mathbb{N}$ is bounded from below $a \in \mathbb{R}$. Assume that $n_j \in \mathbb{N}, j \in \mathbb{N}$ be an increasing sequences of integers such that $\lim_{j\to\infty} \frac{a_{n_j}}{n_j} = a$. Fix $\epsilon > 0$ and assume that $\frac{a_{n_j}}{n_j} \le a + \epsilon$ for some big enough i. Let $k \ge n_j$. Then $k = in_j + r$ for some $i \ge 1$ and $r \in [0, n_j - 1]$. Use subadditivity to conclude $a_k \le a_{in_j} + a_r \le ia_{n_j} + a_r \le in_j(a + \epsilon) + a_r$. This inequality yields $\limsup_{k\to\infty} \frac{a_k}{k} \le a + \epsilon$. Since we can ϵ to be an arbitrary positive small number it follows that

$$a = \liminf_{i\to\infty} \frac{a_i}{i} \le \limsup_{k\to\infty} \frac{a_k}{k} \le a \Rightarrow \lim_{i\to\infty} \frac{a_i}{i} = a.$$

Since $\frac{a_{ij}}{ij} \le \frac{a_j}{j}$ it follows that $a = \lim_{i\to\infty} \frac{a_{ij}}{ij} \le \frac{a_j}{j}$. □

Remark 6.8 *The above lemma holds also in the case that $\frac{a_i}{i}, i \in \mathbb{N}$ is not bounded from below. Then $a = -\infty$.*

Theorem 6.9 *Let $A \in \mathbb{R}^{n\times n}$ and assume that $\rho(A)$ is the spectral radius of A. Then for any $k \in \mathbb{N}$ and $p \in [1, \infty]$ $\rho(A) \le ||A^k||_p^{\frac{1}{k}}$. For $p = 2$ and any $A \in \mathrm{S}_n(\mathbb{R}), k \in \mathbb{N}$ the equality $\rho(A)^k = ||A^k||_2$ holds. Furthermore $\lim_{k\to\infty} ||A^k||_p^{\frac{1}{k}} = \rho(A)$ for any $p \in [1, \infty]$ and $A \in \mathbb{R}^{n\times n}$.*

Proof. Let λ be an eigenvalue of A. Hence $A\mathbf{x} = \lambda\mathbf{x}, \mathbf{x} \ne \mathbf{0}$. Then

$$\lambda^k\mathbf{x} = A^k\mathbf{x} \Rightarrow ||\lambda^k\mathbf{x}||_p = |\lambda|^k||\mathbf{x}||_p = ||A^k\mathbf{x}||_p \le ||A^k||_p||\mathbf{x}||_p \Rightarrow |\lambda|^k \le ||A^k||_p \Rightarrow \rho(A) \le ||A^k||_p^{\frac{1}{k}}.$$

Use Theorem 6.4 and Corollary 6.3 to deduce that for any $A \in \mathrm{S}_n(\mathbb{R})$ we have $\rho(A)^k = ||A^k||_2$. Assume first that $\rho(A) = 0$. Then it is known that A is nilpotent, i.e $A^n = 0 \Rightarrow A^k = 0$ for $k \ge n$. Hence $0 = \rho(A) = \lim_{k\to\infty} ||A^k||_p^{\frac{1}{k}}$. Suppose that $\rho(A)$. Use (6.13) to deduce that the sequence $\log ||A^k||_p, k \in \mathbb{N}$ is a subadditive sequence. Clearly $\log \rho(A) \le \frac{\log ||A^k||_p}{k}$ for any $k \in \mathbb{N}$. Lemma 6.6 implies that $\lim_{k\to\infty} \frac{\log ||A^k||_p}{k} = t \ge \log \rho(A)$. It is left to show that $t = \log \rho(A)$. We will prove the generic case where A is diagonable, i.e. $A = X \operatorname{diag}(\lambda_1, \ldots, \lambda_n)X^{-1}$. Then

$$||A^k||_p^{\frac{1}{k}} = ||X \operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k)X^{-1}||_p^{\frac{1}{k}} \le ||X||_p^{\frac{1}{k}} || \operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k)||_p^{\frac{1}{k}} ||X^{-1}||_p^{\frac{1}{k}} = ||X||_p^{\frac{1}{k}} \rho(A) ||X^{-1}||_p^{\frac{1}{k}}.$$

Let $k \to \infty$ to deduce that $e^t \le \rho(A)$. Hence $e^t = \rho(A)$. □

6.4 Nonnegative symmetric matrices

Denote by $S_n(\mathbb{R}_+)$ the set of all $n \times n$ real symmetric matrices with nonnegative entries. In this subsection we assume that $A \in \mathrm{S}_n(\mathbb{R}_+)$.

Theorem 6.10 *Let $A = (a_{ij})_{i,j=1}^n \ne 0$ be a real symmetric matrix with nonnegative entries. Assume that $G(A)$ is connected. Arrange the eigenvalues of A in a decreasing order (6.8). Then $\lambda_1 > 0$ and $\lambda_1 \ge |\lambda_n|$, i.e. $\lambda_1 = \rho(A)$. Furthermore $\lambda_1 > \lambda_2$, i.e. λ_1 is a simple root of $\det(zI_n - A)$. The corresponding eigenvector $\mathbf{x}_1$ in (6.8) can be chosen to be a vector of length one with positive coordinates. Furthermore $\lambda_1 > |\lambda_n|$ unless $G(A)$ is bipartite. If $G(A)$ is bipartite then $\lambda_n = -\lambda_1$ and $\lambda_1 > |\lambda_i|$ for $i = 2, \ldots, n-1$.*

Proof. For any vector $\mathbf{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$ let $|\mathbf{x}| := (|x_1|, \ldots, |x_n|)^\top$ be a vector with nonnegative coordinates. Note that $\mathbf{x}^\top \mathbf{x} = |\mathbf{x}|^\top |\mathbf{x}|$. As all the entries of A are nonnegative it follows that for any $\mathbf{x} \in \mathbb{R}^n$

$$|\mathbf{x}^\top A \mathbf{x}| = |\sum_{i,j=1}^n a_{ij} x_i x_j| \le \sum_{i,j=1}^n a_{ij} |x_i| |x_j = |\mathbf{x}|^\top A |\mathbf{x}|.$$

Thus in the maximum characterization of (6.9) it is enough to consider all $\mathbf{x}$ whose coordinates are nonnegative. Let $\mathbf{e} = (1, \ldots, 1)^\top$. Then $\lambda_1 \ge \frac{\mathbf{e}^\top A \mathbf{e}}{\mathbf{e}^\top \mathbf{e}} > 0$. The maximum is achieved for some $\mathbf{x}_1, ||\mathbf{x}_1|| = 1$ with nonnegative coordinates. $\mathbf{x}_1$ must be an eigenvector of A corresponding to λ_1, i.e. $A\mathbf{x}_1 = \lambda_1 \mathbf{x}_1$. Note that $(I_n + A)\mathbf{x}_1 = (1 + \lambda)\mathbf{x}_1$. Hence $(I_n + A)^{n-1}\mathbf{x}_1 = (1 + \lambda_1)^{n-1}\mathbf{x}_1$. Since $G(A)$ is connected Lemma 4.4 $(I_n + A)^{n-1}$ has all positive entries. Since $\mathbf{x}_1$ has at least one positive coordinate it follows that $(I_n + A)^{n-1}\mathbf{x}_1$ has all positive coordinates. Thus $\mathbf{x}_1 = (1+\lambda_1)^{-n+1}(I_n+A)^{n-1}\mathbf{x}_1$ has positive coordinates.

Since $|\lambda_i| = |\mathbf{x}_i^\top A \mathbf{x}_i| \le |\mathbf{x}_i|^\top A |\mathbf{x}_i| \le \lambda_1$ it follows that $\lambda_1 \ge |\lambda_i|$ for $i = 2, \ldots, n$. It is left to show that $\lambda_1 > \lambda_i$ for each $i > 1$. Consider the matrix $B = (b_{ij})_{i,j=1}^n = (I_n+A)^{n-1}$ it has eigenvalues $\beta_i := (1 + \lambda_i)^{n-1}$ for $i = 1, \ldots, n$. Clearly $\beta_1 = (1 + \lambda_1)^{n-1} \ge (1 + |\lambda_i|)^{n-1} \ge |(1+\lambda_i)^{n-1}| = |\beta_i|$. it is enough to show that $\beta_1 > \beta_i$ for $i > 1$. We now repeat the arguments for the maximal eigenvalue β_1 of the real symmetric matrix B with positive entries. Let $\mathbf{y}_1$ be a positive eigenvector of B corresponding to $\max_{||\mathbf{y}||=1} \mathbf{y}^\top B \mathbf{y}$. It is known that it is always possible to choose the eigenvector $\mathbf{y}_i$ of B such that $B\mathbf{y}_i = \beta_i \mathbf{y}_i, ||\mathbf{y}_i|| = 1, \mathbf{y}_1^\top \mathbf{y}_i = 0$ for any $i > 1$. Since all the coordinates of $\mathbf{y}_1$ are positive, the condition $\mathbf{y}_1^\top \mathbf{y}_i = 0$ implies that $\mathbf{y}_i$ has positive and negative coordinates for $i > 1$. As each $b_{pq} > 0$ it follows that $|\beta_i| = |\mathbf{y}_i^\top B \mathbf{y}_i| < |\mathbf{y}_i|^\top B |\mathbf{y}_i| \le \beta_1$ for $i > 1$.

Assume now that $G(A)$ is not bipartite. Then $G(A)$ is aperiodic (Corollary 2.19). Then there exist N such that for any $m \ge N$ A^m has positive entries. In particular A^{2N} has positive entries. Thus $G(A^{2N})$ is connected. The eigenvalues of A^{2N} are $\lambda_1^{2N}, \ldots, \lambda_n^{2N}$ are all nonnegative. λ_1^{2N} is the maximal eigenvalue. Hence

$$\lambda_1^{2N} > \lambda_n^{2N} = |\lambda_n|^{2N} \Rightarrow \lambda_1 > |\lambda_n|.$$

Assume that $G(A)$ is bipartite. Then there exists a permutation matrix $Q \in \{0,1\}^{n \times n}$ such that $C = QAQ^\top = \begin{pmatrix} 0 & A_{12} \\ A_{21} & 0 \end{pmatrix}$. (Since $C^\top = C$ it follows that $A_{21} = A_{12}^\top$.) Assume that $A_{12} \in \mathbb{R}^{p \times q}$. Clearly C and A are similar matrices, hence C and A have the same eigenvalues. Note that $C(Q\mathbf{x}_1) = \lambda_1(Q\mathbf{x}_1)$. Let $(Q\mathbf{x}_1)^\top = (\mathbf{u}^\top, \mathbf{v}^\top), \mathbf{u} \in \mathbb{R}^p, \mathbf{v} \in \mathbb{R}^q$ are positive vectors. Then

$$\begin{pmatrix} 0 & A_{12} \\ A_{21} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \lambda_1 \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \Rightarrow \begin{pmatrix} 0 & A_{12} \\ A_{21} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ -\mathbf{v} \end{pmatrix} = -\lambda_1 \begin{pmatrix} \mathbf{u} \\ -\mathbf{v} \end{pmatrix}.$$

Hence $-\lambda_n = \lambda_n$. Consider $C^2 = \operatorname{diag}(A_{12}A_{21}, A_{21}A_{12})$. Since $G(C^2)$ is isomorphic to $G(A^2)$ it follows that Corollary 2.19 that $A_{12}A_{21}, A_{21}A_{12}$ are irreducible and aperiodic. Note that

$$A_{12}A_{21}\mathbf{u} = \lambda_1^2 \mathbf{u}, \quad A_{21}A_{12}\mathbf{v} = \lambda_1^2 \mathbf{v}.$$

As the eigenvalues of C^2 are $\lambda_1^2, \ldots, \lambda_n^2$, which are nonnegative and all the eigenvalues of $A_{12}A_{21}, A_{21}A_{12}$ different from $\lambda_1^2 = \lambda_n^2$ are strictly less than λ_1^2, we deduce that $|\lambda_i| < \lambda_1$ for $i = 2, \ldots, n-1$. □

The arguments of the proof of this theorem that yield that $\lambda_n = -\lambda_1$ in the case $G(A)$ is bipartite imply:

Corollary 6.11 *Let $A \in \mathrm{S}_n(\mathbb{R})$ and assume that $G(A)$ is bipartite. Then the first $\lfloor \frac{n}{2} \rfloor$ eigenvalues of A are nonnegative and the last $\lfloor \frac{n}{2} \rfloor$ are nonpositive. Furthermore $\lambda_{n-i+1} = -\lambda_i$ for $i = 1, \ldots, n$.*

Theorem 6.12 *Let $A \in \mathrm{S}_n(\mathbb{R}_+)$ and assume that A is irreducible. Let $A\mathbf{x}_1 = \lambda_1\mathbf{x}_1$ where $\mathbf{x}_1$ is a positive vector of norm 1. Then*

$$||\lambda_1^{-m}A^m - \mathbf{x}_1^\top\mathbf{x}_1|| = r^m, \ \texttt{where}\ r := \max_{i\in[2,n]}\frac{|\lambda_i|}{\lambda_1}, \ \texttt{for any}\ m \in \mathbb{N}. \tag{6.14}$$

If $G(A)$ is not bipartite then $\lim_{m\to\infty}\lambda_1^{-m}A^m = \mathbf{x}_1^\top\mathbf{x}_1$. If $G(A)$ is bipartite then A^2 decomposes as to a direct sum of $A_{12}A_{21} \oplus A_{12}A_{12}$, where each summand is an irreducible aperiodic symmetric matrix. In particular $\lim_{m\to\infty}\lambda_1^{-2m}A^{2m}$ exists and is a direct sum of two symmetric rank one matrices, each one having one nonzero eigenvalue equal to 1.

Proof. Recall that $A = X\,\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)X^\top$ for an orthogonal X, whose first column is $\mathbf{x}_1$. Observe next that $\mathbf{x}_1\mathbf{x}_1^\top = X\,\mathrm{diag}(1, 0, \ldots, 0)X^\top$. Use Theorem 6.9 to deduce

$$||\lambda_1^{-m}A^m - \mathbf{x}_1^\top\mathbf{x}_1 = ||X\,\mathrm{diag}(1, \frac{\lambda_2^m}{\lambda_1^n}, \ldots, \frac{\lambda_n^m}{\lambda_1^n})X^\top - X\,\mathrm{diag}(1, 0, \ldots, 0)X^\top|| =$$
$$||X\,\mathrm{diag}(0, \frac{\lambda_2^m}{\lambda_1^n}, \ldots, \frac{\lambda_n^m}{\lambda_1^n})X^\top|| = r^m.$$

This proves (6.14). If $G(A)$ is connected and not bipartite then $r < 1$. Hence

$$\lim_{k\to\infty}\lambda_1^{-m}A^m = \mathbf{x}_1\mathbf{x}_1^\top = X\,\mathrm{diag}(1, 0, \ldots, 0)X^\top, \tag{6.15}$$

and the theorem follows in this case. The second case follows from the proof of Theorem 6.10. □

Theorem 6.13 *Let $P \in \mathrm{S}_n(\mathbb{R}_+)$ be a stochastic matrix. Let $G(P) = (\langle n\rangle, E)$ be the induced undirected graph by P.*
1. Assume that $G(P)$ is connected. Then $\frac{1}{\sqrt{n}}\mathbf{1}, \mathbf{1} := (1, \ldots, 1)^\top \in \mathbb{R}^n$ is the unique positive P-eigenvector of length one corresponding to the eigenvalue $\lambda_1 = 1$. All other eigenvalue of P are real and $\lambda_i \in (-1, 1]$ for $i = 2, \ldots, n$. P has unique stationary distribution $\mu = \frac{1}{n}(1, \ldots, 1) = \frac{1}{n}\mathbf{1}^\top$, called equidistribution.
(a) Suppose that $G(P)$ is not bipartite. Then $\lambda_i \in (-1, 1)$ for $i = 1, \ldots, n$,
$r := \max(|\lambda_2|, |\lambda_n|) < 1$ and

$$||P^m - \mathbf{1}(\frac{1}{n}\mathbf{1}^\top)| \le r^m, \ ||\mu^{(0)}P^m - \frac{1}{n}\mathbf{1}^\top|| \le ||\mu^{(0)}||r^m \le r^m, \ \texttt{for any}\ m \in \mathbb{N},$$
$$\lim_{m\to\infty}P^m = \mathbf{1}(\frac{1}{n}\mathbf{1}^\top) \ \lim_{m\to\infty}\mu^{(0)}P^m = \mu = \frac{1}{n}\mathbf{1}^\top \ \texttt{for any}\ \mu^{(0)} \in \Pi_n. \tag{6.16}$$

(b) Suppose that $G(P)$ is bipartite. Then $\lambda_n = -1$ and $\lambda_i \in (-1, 1)$ for $i = 1, \ldots, n-1$. P^2 reduces to a direct sum of two stochastic matrices $P_1 \oplus P_2$, where each $G(P_i) = (V_i, E_i)$ is connected and not bipartite. For each $\mu^{(0)} \in \Pi_n$ $\lim_{m\to\infty}\mu^{(0)}P^{2m} = a\mu_1 \oplus (1-a)\mu_2$, where μ_i is the equidistribution corresponding to P_i and $a \in [0, 1]$ is $\Pr(X_0 \in V_1)$, i.e. the sum of the coordinates on $\mu^{(0)}$ corresponding to V_1.
2. Assume that $G(P)$ is not connected and let $G(V_1), \ldots, G(V_k)$ be the connected components of $G(P)$. Then each $P(V_i)$ is a symmetric and irreducible stochastic matrix. Furthermore $\lim_{m\to\infty}P^{2m} = P_{2,\infty}$ is symmetric stochastic matrix, which has eigenvalues 1, of multiplicity $m \ge k$, and all other eigenvalues are equal to zero. Each connected component contributes is either 1 to m, if $G(V_i)$ is not bipartite, or 2 if $G(V_i)$ is bipartite. Furthermore $P_{2,\infty}^2 = P_{2,\infty}$. If each connected component is not bipartite then $m = k$ and $\lim_{m\to\infty}P^m = P_{2,\infty} = \oplus_{i=1}^k\mathbf{1}\mu_i(V)$, where $\mu_i(V)$ is the unique stationary measure corresponding to $P(V_i)$ for $i = \ldots, k$.

Proof. Using the previous theorems it is enough to prove the case (1a). The inequality $||P^m - \mathbf{1}(\frac{1}{n}\mathbf{1}^\top)|| \leq r^m$ follows from (6.14). Furthermore, since $\mu = \mu^{(0)}\mathbf{1}(\frac{1}{n}\mathbf{1}^\top)$ it follows

$$||\mu^{(0)}P^m - \frac{1}{n}\mathbf{1}^\top|| = ||\mu^{(0)}(P^m - \mathbf{1}(\frac{1}{n}\mathbf{1}^\top))|| \leq ||\mu^{(0)}||P^m - \mathbf{1}(\frac{1}{n}\mathbf{1}^\top)|| = ||\mu^{(0)}||r^m.$$

Clearly $||\mu^{(0)}|| \leq \sqrt{||\mu^{(0)}||_1} = 1$. This proves all the inequalities in (6.16). The equalities parts of (6.16) follow easily.

6.5 Reversible Markov Chains

A stochastic matrix $P = (p_{ij}) \in \mathbb{R}_+^{n\times n}$ is *reversible* with respect to the distribution $\pi = (\pi_1, \ldots, \pi_n) \in \Pi_n$ if

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \texttt{for all } i, j = 1, \ldots, n. \tag{6.17}$$

Lemma 6.14 *Let $P \in \mathbb{R}_+^{n\times n}$ be a stochastic matrix reversible with respect to $\pi \in \Pi_n$. Then π is a stationary distribution of P. Let $V = \{i_1, \ldots, i_k\} = \operatorname{supp}\pi \subset \langle n\rangle$ be the set of vertices of $G(P)$ where $\pi_j > 0$. Then $P(V)$ is stochastic and $\pi(V) \in \Pi_k$ is a stationary distribution of $P(V)$. Let $D = \operatorname{diag}(\sqrt{\pi_1}, \ldots, \sqrt{\pi_n})$. Then $D(V)$ is a diagonal matrix such that its all diagonal are positive. Furthermore $T := D(V)PD(V)^{-1}$ is a nonnegative symmetric matrix, which is diagonally similar to $P(V)$. Hence $P(V)$ has only real eigenvalues and each eigenvalue of P is geometrically simple.*

Proof. Fix $j \in [1, n]$ in (6.17) and sum on $i = 1, \ldots, n$. Since P is stochastic we obtain $\pi P = \pi$, i.e. π is a stationary distribution. Let $V^c := \langle n\rangle \backslash V$. Suppose first that $V^c \neq \emptyset$. Assume in (6.17) that $i \in V$ and $j \in V^c$. Then $p_{ij} = 0$. Hence there are no edges from V to V^c and $P(V)$ is a stochastic submatrix of P. If $V = \langle n\rangle$ then $P(V) = P$ is stochastic. Furthermore $\pi(V)$ is the stationary distribution of $P(V)$.

To show the rest of the lemma, one can assume for simplicity of the argument that $V = \langle n\rangle$. Let $T = DPD^{-1}$. Then is straightforward to show that (6.17) is equivalent to T being a symmetric matrix. Thus P is diagonally similar to a symmetric matrix with nonnegative entries T. Let $\lambda_1 \geq \ldots \geq \lambda_n$. So $T = X\operatorname{diag}(\lambda_1, \ldots, \lambda_n)X^\top$ and $P = D^{-1}X\operatorname{diag}(\lambda_1, \ldots, \lambda_n)X^\top D$. Thus $1 = \lambda_1 \geq \ldots \geq \lambda_n$ are the eigenvalues of P and each eigenvalue of P is geometrically simple. □

Theorem 6.15 *Let $\pi = (\pi_1, \ldots, \pi_n)$ be a probability vector with $\pi_i > 0$ for $i = 1, \ldots, n$. Let $P \in \mathbb{R}_+^{n\times n}$ be a stochastic matrix reversible with respect to π. Let $D = \operatorname{diag}(\sqrt{\pi_1}, \ldots, \sqrt{\pi_n})$. Then $T := DPD^{-1}$ is a nonnegative symmetric matrix, which is diagonally similar to P. Then $G := G(P) = (\langle n\rangle, E)$ is a reversible digraph: $(i, j) \in E \iff (j, i) \in E$. Let $\langle n\rangle = \cup_{i=1}^k V_i$ corresponds to the decomposition of G to its connected components $G(V_1), \ldots, G(V_k)$. Then $P := \oplus_{i=1}^k P(V_i)$, where each $P(V_i)$ is an irreducible stochastic matrix reversible with respect to the distribution $\mu_i := \frac{1}{\sum_{j\in V_i}\pi_j}\pi(V_i)$ for $i = 1, \ldots, k$.*

Assume now that $G(P)$ is connected and let $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n$.

1. Assume that $G(P)$ is not bipartite, i.e. $-\lambda_n < \lambda_1$. Then $r := \max_{i=2,\ldots,n}|\lambda_i| < 1$ and

$$||D(P^m - \mathbf{1}\pi)D^{-1}|| \leq r^m,\ ||(\mu^{(m)} - \pi)D^{-1}|| \leq ||\mu^{(0)}D^{-1}||r^m \Rightarrow \lim_{m\to\infty} P^m = \mathbf{1}\pi. \tag{6.18}$$

2. Assume that $G(P)$ is bipartite where $\langle n\rangle = V \cup V^c$ and $E \subset V \times V^c \cup V^c \times V$. Then $P^2 = Q(V)\oplus Q(V^c)$, where $Q(V^c), Q(V^c)$ are irreducible stochastic reversible matrices with respect to $\mu_1 := \frac{1}{\sum_{i\in V}\pi_i}\pi(V), \mu_2 := \frac{1}{\sum_{i\in V^c}\pi_i}\pi(V^c)$ respectively. Furthermore $G(Q(V)), G(Q(V^c))$ are connected not bipartite matrices. Hence

$$\lim_{m\to\infty} P^{2m} = \mathbf{1}\mu_1 \oplus \mathbf{1}\mu_2, \tag{6.19}$$

and the convergence geometric in $r := \max_{i=2,\ldots,n-1}\lambda_i^2$. That is $||D(P^{2m} - \mathbf{1}\mu_1 \oplus \mathbf{1}\mu_2)D^{-1}|| \leq r^m$.

Proof. The first part of the theorem follows straightforward from Lemma 6.14. The second part follows from the fact that $T := DPD^{-1} \in S_n(\mathbb{R}_+)$, Theorem 6.12 and the arguments of the proof of Theorem 6.13. □

Note that Theorem 5.2 follows form the above theorem if P is a reversible stochastic matrix with respect to a distribution $\pi = (\pi_1, \dots, \pi_n)$, where $\pi_i > 0$ for $i = 1, \dots, n$. Moreover of $P = P^\top$ is stochastic, then P is reversible with respect to the uniform distribution $\pi = \frac{1}{n}\mathbf{1}$.

We now explain the terminology reversible Markov chain. Assume that (6.17) holds. Then by Lemma 6.17 π is a stationary distribution. Let $X_0, \dots,$ be the Markov chain associated with P, with the stationary distribution π. Then the condition (6.17) is equivalent to $\Pr(X_{m-1} = i, X_m = j) = \Pr(X_{m-1} = j, X_m = i)$ for each $i, j \in \langle n \rangle$. That is we can reverse the random walk in time with the same probability. More generally

$$\Pr(X_0 = i_0, X_i = i_1, \dots, X_m = i_m) = \Pr(X_0 = i_m, X_{i_1} = i_{m-1}, \dots, X_m = i_0).$$

6.6 Markov chains associated with digraphs

Let $G = (\langle n \rangle, E)$ be a digraph where for which vertex i the outdegree $\deg_{\text{out}}(i) \geq 1$ for each $i = 1. \dots, n$. With such G we associate the following stochastic matrix

$$P(G) := P = (p_{ij})_{i,j=1}^n \in \mathbb{R}_+^{n \times n}, \quad p_{ij} = 0 \ \texttt{if} \ (i,j) \notin E, \quad p_{ij} = \frac{1}{\deg_{\text{out}}(i)} \ \texttt{if} \ (i,j) \in E. \tag{6.20}$$

That is $G(P) = G$ and from each i the probability to go any vertex j, such that i is connected to j, is constant, i.e. equal to the 1 divided by number of vertices which can be reached from i.

Assume G is reversible, i.e $(i,j) \in E \iff (j,i) \in E$. Then $\deg_{\text{out}}(i) = \deg_{\text{in}}(i), i = 1, \dots, n$. $\pi_i = \frac{\deg_{\text{out}}(i)}{\#E}, i = 1, \dots, n$, where $\#E = \sum_{i=1}^n \deg_{\text{out}}(i)$ is the number of directed edges in G. A straightforward calculation shows that P is a reversible with respect to the distribution $\pi = (\pi_1, \dots, \pi_n)$. For this Markov chain we can use Theorem 6.15.

The problem with $P(G)$, even if it is irreducible it may be periodic. To avoid this possibility, sometimes one uses instead of stochastic matrix $G(P)$ the stochastic matrix $P(G, a) := aI_n + (1-a)P(G)$ for some $a \in (0,1)$. This guarantees that the loop $(i,i) \in G(P,a)$ for any $a \in (0,1)$. Then the restriction of $P(G,a)$ to each terminal $\emptyset \neq V \subset \langle n \rangle$, such that $\{V\}$ is a terminal vertex in the reduced graph of G_{rdc}, is an aperiodic stochastic submatrix of $P(G,a)$. (This trick is used also in the Google search engine!) This guarantees that $\lim_{m \to \infty} P(G,a)^m = Q$, where Q is some special stochastic matrix associated with G independent of $a \in (0,1)$.

In the case that $a = 0.5$ then $P(G, 0.5)$ represents *lazy* random walk on the graph G, since with a probability at least 0.5 the particle stays at vertex i at time $m+1$ if it was at time m at the vertex i.

7 Perron-Frobenius theorem

Theorem 7.1 *Let $0 \neq A \in \mathbb{R}_+^{n \times n}$ be an irreducible matrix, i.e. $G(A)$ is strongly connected. Then $\rho(A)$ is an eigenvalue of A , which is a simple root of $\det(zI - A)$. The eigenspace $E(\rho(A))$ is spanned by a positive vector $\mathbf{v} > 0$. Let $\mathbf{u} > 0$ be a left positive eigenvector of A corresponding to $\rho(A)$: $\mathbf{u}^\top A = \rho(A)\mathbf{u}^\top$. Assume the normalization condition $\mathbf{u}^\top \mathbf{v} = 1$. Then eigenvector $\mathbf{z} \in \mathbb{C}^n$ corresponding to an eigenvalue $\lambda \neq \rho(A)$ of A satisfies $\mathbf{u}^\top \mathbf{z} = 0$.*

1. Assume that $A = (a_{ij})_{i.j=1}^n$ is aperiodic, i.e. $G(A)$ is aperiodic, then then $\rho(A) = \lambda_1 > |\lambda_1| \geq \dots \geq |\lambda_n| \geq 0$ and $\lim_{m \to \infty} \lambda_1^{-m} A^m = \mathbf{v}\mathbf{u}^\top$. This convergence is geometric in

$r := \max_{i\in[2,n]} \frac{|\lambda_i|}{|\lambda_1|}$, *i.e.* $||\lambda_1^{-m}A^m - \mathbf{v}\mathbf{u}^\top||_p \le K_p(A)r^m$, *for some constant* $K_p(A)$ *depending on* A *and* $p \in [1,\infty]$. *Furthermore*

$$\rho(A) = \max_{\mathbf{x}=(x_1,\ldots,x_n)^\top>\mathbf{0}} \min_{i\in[1,n]} \frac{\sum_{j=1}^n a_{ij}x_j}{x_i} = \min_{\mathbf{x}=(x_1,\ldots,x_n)^\top>\mathbf{0}} \max_{i\in[1,n]} \frac{\sum_{j=1}^n a_{ij}x_j}{x_i}. \tag{7.1}$$

(This is Wielandt's characterization.)
2. Assume that A *is* $2 \le p$*-periodic, i.e.* $G(A)$ *is* $2 \le p$*-periodic. Then* A *has exactly* p *algebraically simple eigenvalues on the circle* $|z| = \rho(A)$*:* $\lambda_i = \rho(A)e^{\frac{(i-1)\sqrt{-1}}{p}}$ *for* $i = 1,\ldots,p$. *All other eigenvalues of* A *satisfy* $|\lambda| < \rho(A)$*:* $\rho(A) = \lambda_1 = |\lambda_2| = \ldots = |\lambda_p| > |\lambda_{p+1}| \ge \ldots \ge |\lambda_n|$. *Moreover the spectrum of* A *is invariant under the multiplication by* $e^{\frac{\sqrt{-1}}{p}}$. *(Hence by any* $p-th$ *root of unity). There exists a permutation matrix* P *such that* $PAP^\top$ *is of the form (4.4). In particular the eigenvectors corresponding to* λ_i *can be easily obtained from* $\mathbf{v}$ *for* $i = 2,\ldots,p$. *(Similarly to the situation described in the proof of Theorem 6.10 in the case* $G(A)$ *is bipartite.) The matrix* $A^p = \oplus_{i=1}^p A_i$, *where* A_i *satisfies the conditions (1) f the theorem and* $\rho(A_i) = \rho(A)^p$ *for* $i = 1,\ldots,p$. *In particular* $\lambda_1^{-m}A^{pm}$ *converges as* $m \to \infty$ *to rank* p *nonnegative matrix, which is a direct sum of rank one positive matrices.*

Historical remarks: The above theorem, (with small variations), is called the Perron-Frobenius theorem. For matrices with positive entries this result is due to Perron 1907. For nonnegative irreducible matrices this result is due to Frobenius 1908, 1909 and 1912. The minimax characterization (7.1) is due to Wielandt 1950.

Theorem 7.2 *Let* $A \in \mathbb{R}_+^{n\times n}$. *Let* $G := G(A)$ *and denote by* G_{rdc} *the reduced graph of* G, *which is acyclic. Assume that* G *has* k *vertices. (*$k = 1$ *corresponds to the case* G *is strongly connected.)*

Then A *is permutationally similar to a nonnegative matrix of the form of the form (4.5). Each* A_{ii} *is irreducible,* $\det(zI - A) = \prod_{i=1}^k \det(zI - A_{ii})$ *and* $\text{spec}\,(A) = \cup_{i=1}^k \text{spec}\,(A_{ii})$. *Hence* $\rho(A)$ *is an eigenvalue of* A. *There exists a nonzero nonnegative eigenvector* $\mathbf{v}$ *corresponding to* $\rho(A)$*:* $A\mathbf{v} = \rho(A)\mathbf{v}$. *Furthermore:*

$$\lim_{m\to\infty} \left(\mathbf{1}^\top A^m \mathbf{1}\right)^{\frac{1}{m}} = \rho(A) = \limsup_{m\to\infty} \left(\operatorname{tr} A^m\right)^{\frac{1}{m}}. \tag{7.2}$$

Call a vertex $\{V_i\}$ *in* G_{rdc}, *corresponding to an irreducible matrix* A_{ii}, *singular if* $\rho(A_{ii}) = \rho(A)$, $(\det(\rho(A)I - A_{ii}) = 0)$. *Let* j *be the maximal number of singular vertices in all possible paths in* G_{rdc}. *(*$j = 1$ *may corresponds to a path of length* 0.*) Then* j *is the maximal size of a Jordan block in* A *corresponding to* $\rho(A)$. *Furthermore, the size of the Jordan block of any eigenvalue* λ *of* A *satisfying* $|\lambda| = \rho(A)$ *does not exceed the above* j.

Proof. Consider $A_k = A + \frac{1}{k}\mathbf{1}\mathbf{1}^\top$. So each A_k has positive entries and $\lim_{k\to\infty} A_k = A$. Now use the fact that $\lim_{k\to\infty} \det(zI_n - A_k) = \det(zI_n - A)$. Hence the eigenvalues of A_k, counted with their multiplicities converge to the eigenvalues of A. So $\lim_{k\to\infty} \rho(A_k) = \rho(A)$ and $\rho(A) \in \text{spec}\,(A)$. Let $A_k\mathbf{v}_k = \rho(A_k)\mathbf{v}_k$ where $\mathbf{v}_k > 0$ and $||\mathbf{v}_k|| = 1$. By taking a convergent subsequence $\mathbf{v}_{k_q} \to \mathbf{v}$ we deduce that $\mathbf{v} \ge \mathbf{0}, ||\mathbf{v}|| = 1$ and $A\mathbf{v} = \rho(A)\mathbf{v}$. Clearly $\text{spec}\,(A) = \cup_{i=1}^k \text{spec}\,(A_{ii})$.

Observe that $||A^m||_1 \le \mathbf{1}^\top A^m \mathbf{1} \le n||A^m||_1$. Use Theorem 6.9 for $p = 1$ to obtain the first equality in (7.2). Since $\mathbf{1}^\top A^m \mathbf{1} \ge \operatorname{tr} A^m$ we deduce from the first equality in (7.2) that $\rho(A) \ge \limsup_{m\to\infty} \left(\operatorname{tr} A^m\right)^{\frac{1}{m}}$. Use the fact that $\text{spec}\,(A) = \cup_{i=1}^k \text{spec}\,(A_{ii})$ and use Perron-Frobenius theorem to deduce the second equality in (7.2).

The last claim about the maximal size of the Jordan block corresponding to $\rho(A)$ is due to Rothblum [19]. The claim that the maximal size of the Jordan block corresponding to am eigenvalue λ on the circle $|z| = \rho(A)$ is well known to the experts. See for example [9] for the proofs of these claims. □

The proof of Theorem 5.2 follows mainly from Theorem 7.1, Theorem 7.2 and Proposition 5.1.

8 Symbolic dynamics and walks on graphs

8.1 Introduction

View $\langle n\rangle = \{1,\dots,n\}$ as an alphabet on n letters. Consider all the words of this alphabet of length m: $\mathbf{a}_m := (a_1, a_2, \dots, a_m)$, where $a_i \in \langle n\rangle$. Any word $\mathbf{a}_m$ can be viewed as a walk of length m on the complete digraph $KD_n := (\langle n\rangle, \langle n\rangle \times \langle n\rangle)$. That is, the walk is given by $a_1 \to a_2 \to \dots \to a_m$. Clearly the number of all walks of length m is n^m. In *electrical engineering* all words $\mathbf{a}_m$ are viewed as a transmission of information of length m. It is called an unconstrained, (free), *channel*. The $m-th$ capacity of free channel is $\log n = \frac{\log n^m}{m}$. In information theory, one usually considers logarithms on basis 2. So the capacity of the free channel is $\log_2 n$. In mathematics and physics one usually uses the logarithms on the natural basis e, which is here denoted simply by $\log x$.

In many cases one has some local restrictions on the form of a word $\mathbf{a}_m$ that can be transmitted. This restrictions can be always translated to the assumption that $\mathbf{a}_m$ is a walk on some graph $G = (\langle n\rangle, E)$.

Consider the following example. Let $G := (\langle 2\rangle, \{(1,2),(2,1),(2,2)\})$. That G is a reversible digraph such that from the state 1 one can go to the state 2, and from the state 2 one can go to the states 1 and 2. In this example it is common to replace 2 by 0. Thus $\mathbf{a}_m = (a_1, a_2, \dots, a_m)$ is a signal of length m, consisting of 0 and 1, such that no two 1 are adjacent. Let l_m be the number of words $\mathbf{a}_m$ that satisfy the above restriction. Equivalently, l_m is the number of walk on G of length m. Clearly $l_1 = 2, l_2 = 3$. We claim that $l_1, l_2, l_3, \dots$ is a Fibonacci sequence:

$$l_m = l_{m-2} + l_{m-1}, \quad m = 3, \dots \tag{8.1}$$

Indeed, consider the word $\mathbf{a}_m = (a_1, a_2, \dots, a_m)$. Suppose first that $a_m = 0$. Then $(a_1, \dots, a_{m-1})$ is any allowable word of length $m-1$. Thus we have l_{m-1} words of length m which end in $0 = a_m$. Now suppose that $a_m = 1$. Then $a_{m-1} = 0$ and $(a_1, \dots, a_{m-2})$ is any allowable word of length $m-2$. This proves (8.1).

Try a solution of (8.1) of the form $l_m = t^m$. Then t satisfies quadratic equation $t^2 = t+1$ which has two solutions $t_1 = \frac{1+\sqrt{5}}{2} > 0 > t_2 = \frac{1-\sqrt{5}}{2}$. The number t_1 is called the *golden ration*. The general solution of (8.1) is given by $l_m = a_1 t_1^m + a_2 t_2^m$. The initial conditions $l_1 = 2, l_2 = 3$ yield that

$$l_m = \big(\frac{5+3\sqrt{5}}{10}\big)\big(\frac{1+\sqrt{5}}{2}\big)^m + \big(\frac{5-3\sqrt{5}}{10}\big)\big(\frac{1-\sqrt{5}}{2}\big)^m, \quad m = 1, 2, \dots \tag{8.2}$$

Definition 8.1 *Let $G = (V, E)$ be an undirected graph. A configuration $\phi : V \to \{0,1\}$, i.e. an assignment of 0 or 1 to each vertex of V, is called allowable if $\phi(u) + \phi(v) < 2$ for any two adjacent vertices $u, v \in V$. (That is, one can not assign to any pair of adjacent vertices u, v values $\phi(u) = \phi(v) = 1$.) The set of allowable configurations $\Phi \subset V^{\{0,1\}}$ ia called in physics the hard core model. An allowable configuration $\phi \in \Phi$ is called a hard core configuration.*

Let $C_m = (\langle m\rangle, E_m)$ be an undirected path on m vertices: $1 - 2 - \dots - m$. ($E_m = ((1,2),(2,3),\dots,(m-1,m))$.) The the hard core configuration on C_m is a word of length m in $\{0,1\}$, where two 1 are not adjacent. Let Φ_m be the set of all hard core configurations on C_m. Then $\#\Phi_m$, (the number of elements in Φ_m), is equal to l_m. Introduce a uniform probability on Φ_m: $\Pr(\phi) = \frac{1}{l_m}$ for any $\phi \in \Phi_m$. Let $X_m : \Phi_m \to \langle m\rangle$ be the random variable $X_m(\phi) = \sum_{i=1}^m \phi(i)$, which assigns to hard cover configuration $\phi \in \Phi_m$ the number of 1 in this configuration.

Let us compute $\mathrm{E}(X_m)$. Denote by $s_m := \sum_{\phi\in\Phi_m} X_m(\phi)$. s_m is the number of 1 all hard core configurations on C_m. Note that $s_1 = 1, s_2 = 2, s_3 = 5, s_4 = 10$. It can be shown that we have the following recursive relation

$$s_m = s_{m-1} + s_{m-2} + l_{m-2}, \quad m = 3, \dots, \tag{8.3}$$

where $l_j, j = 1, \ldots,$ is the Fibonacci sequence defined in (8.1). (That is a homework problem!). It is known that the general solution of (8.3), under the assumption that l_m satisfies (8.1), is of the form

$$s_m = (mb_1 + b_2)\big(\frac{1+\sqrt{5}}{2}\big)^m + (mb_3 + b_4)\big(\frac{1-\sqrt{5}}{2}\big)^m, \quad m = 1, \ldots \tag{8.4}$$

The values of b_1, b_2, b_3, b_4 are completely determined by s_1, s_2, s_3, s_4. One can show that $b_1 > 0$. Thus $\mathrm{E}(X_m) = \frac{s_m}{l_m}$. From these results we deduce

$$\lim_{m\to\infty} \frac{\log l_m}{m} = \log\big(\frac{1+\sqrt{5}}{2}\big), \tag{8.5}$$

$$\lim_{m\to\infty} \frac{\mathrm{E}(X_m)}{m} = \frac{10b_1}{3+\sqrt{5}}. \tag{8.6}$$

(This is a homework problem too!)

Finally consider the following problem. For $t \in \mathbb{Z}_+$ the integer $f(t,m) := l_m \Pr(X_m = t)$ is the number of all configurations in Φ_m with t 1's. Note that for $t > \lceil \frac{m}{2} \rceil$ $f(t,m) = 0$.. Fix $p \in [0, 0.5)$ and consider the sequence of integers $t_m \in \mathbb{Z}_+, m \in \mathbb{N}$ such that $\lim_{m\to\infty} \frac{t_m}{m}$. Is it true that the sequence $\frac{\log f(t_m,m)}{m}, m \in \mathbb{N}$ converges to some function $h(p)$, which depends only on p and not on a particular sequence? We will show that the answer to this problem is *positive*.

The next subsections give generalizations to the problems posed by the hard core model on C_m, where m tends to ∞.

8.2 Shannon capacity of a channel

Consider an alphabet of n letters denoted by $\langle n \rangle$. Let $G = (\langle n \rangle, E)$ be a digraph which contains a cycle. A word $\mathbf{a}_m = (a_1, a_2, \ldots, a_m)$ is called *allowable*, or G-allowable, if the letter j can follow the letter i, i.e. $(i,j) \in G$. Equivalently the word $\mathbf{a}_m$ describes a walk on G, where a_q is the vertex location on the G at time q. (Time starts at $q = 1$ and then is equal $2, 3, \ldots$) Let $\mathrm{W}(m)$ be the set of all allowable words of length m. The assumption that G contains a cycle is equivalent to the assumption that $\mathrm{W}(m) \neq \emptyset$ for each $m \in \mathbb{N}$.

Theorem 8.2 *Let $n \in \mathbb{N}$ and denote by $\langle n \rangle$ an alphabet on n letters. Let $G = (\langle n \rangle, E)$ be a digraph which contains a cycle. Denote by $A = A(G)$ the representation matrix of G and let $\rho(A) = \rho(G)$ be the spectral radius of A, (and hence of G). Let $\mathrm{W}(m)$ be the set of G-allowable words of length m. Denote $l_m := \#W(m), m \in \mathbb{N}$. Then $l_m = \mathbf{1}^\top A^m \mathbf{1}$ and $\log l_m, m \in \mathbb{N}$ is a nonnegative subadditive sequence. Hence $h(G) := \lim_{m\to\infty} \frac{\log l_m}{m}$ exists, and is called the Shannon capacity of the channel, given by G. Furthermore $h(G) = \log \rho(G)$ and $h(G) \leq \frac{\log l_m}{m}$ for any $m \in \mathbb{N}$.*

Proof. Recall that if $A^m = (a_{ij}^{(m)})_{i,j=1}^n$ then $a_{ij}^{(m)}$ is the number of walks from i to j in m steps. (Lemma 4.1.) Hence $l_m = \sum_{i,j=1}^n a_{ij}^{(m)} = \mathbf{1}^\top A \mathbf{1}$. Since G has a cycle $l_m \geq 1$. We next observe that $l_{p+q} \leq l_p l_q$ for any $p, q \in \mathbb{N}$. Indeed and word of length $p+q$ $a_{p+q} = (a_1, \ldots, a_{p+q}) \in \mathrm{W}(p+q)$ is a superposition of two words $\mathbf{b} = (a_1, \ldots, a_p) \in \mathrm{W}(p), \mathbf{c}_p = (a_{p+1}, \ldots, a_{p+q}) \in \mathrm{W}(q)$. In general, a superposition of two words $\mathbf{b} = (b_1, \ldots, b_p) \in \mathrm{W}(p), \mathbf{c} = (c_1, \ldots, c_q)$ to $\mathbf{a} = (b_1, \ldots, b_p, c_1, \ldots, c_q)$ does not have to be in $\mathrm{W}(p+q)$. (It depends if $(b_p, c_1) \in E$ or not.) Hence $l_{p+q} \leq l_p l_q$. Thus the sequence $\log l_m, m \in \mathbb{N}$ is a nonnegative subadditive sequence. Lemma 6.7 yields that the sequence $\frac{\log l_m}{m}$ converges to a (nonnegative) limit denoted by $h(G)$. Furthermore $h(G) \leq \frac{\log l_m}{m}$ for any $m \in \mathbb{N}$. The equality $l_m = \mathbf{1}^\top A^m \mathbf{1}$ and (7.2) yield that $h(G) = \log \rho(A)$. □

Denote by $\langle n \rangle^{\mathbb{N}}$ the set of all mapping $\mathbf{a} : \mathbb{N} \to \langle n \rangle$. That is $\mathbf{a}$ can be identified with an infinite sequence $\mathbf{a} = (a_1, a_2, \ldots)$ where $a_i \in \langle n \rangle$ for any $i \in \mathbb{N}$. One can view $\mathbf{a}$ as an infinite

word on the alphabet $\langle n \rangle$. $\mathbf{a}$ is called periodic if there exists $q \in \mathbb{N}$ such that $a_{i+q} = a_i$ for all $i \in \mathbb{N}$. A word satisfying this condition is called *q-periodic*. Assume that $\mathbf{a}$ is periodic. The smallest $q \in \mathbb{N}$ for which $\mathbf{a}$ is q-periodic is called the period of $\mathbf{a}$. Thus if the period of $\mathbf{a}$ is p then $\mathbf{a}$ is q-periodic iff p divides q.

Let $G = (\langle n \rangle, E)$ be a digraph that contains a loop. Denote by

$$\langle n \rangle^{\mathbb{N}}(G) := \{\mathbf{a} = (a_1, \ldots) \in \langle n \rangle^{\mathbb{N}} : \ (a_i, a_{i+1}) \in E, i = 1, 2, \ldots\}.$$

Thus $\mathbf{a} \in \langle n \rangle^{\mathbb{N}}(G)$ can be considered an infinite allowable word or an infinite walk on G. Note $\mathbf{a} \in \langle n \rangle^{\mathbb{N}}(G)$ is q-periodic then $\mathbf{a}_q := (a_1, \ldots, a_q) \in \mathrm{W}(q)$ and $(a_q, a_1) \in E$. The set of all q-allowable words $\mathbf{a}_q \mathbf{n} \mathrm{W}(q)$ satisfying the condition $(a_q, a_1) \in E$ is denoted by $\mathrm{W}_{\mathrm{per}}(q)$. It is possible that $\mathrm{W}_{\mathrm{per}}(q)$ is empty for some q. (For example assume that G consists exactly of one cycle.) Thus $\mathbf{a}_q \in \mathrm{W}_{\mathrm{per}}(q)$ iff $\mathbf{a}_q = (a_1, \ldots, a_q)$ can be extended to a q-periodic word $\mathbf{a} \in \langle n \rangle^{\mathbb{N}}(G)$. Equivalently $\mathrm{W}_{\mathrm{per}}(q)$ can be identified with all $\mathbf{a}_{q+1} = (a_1, \ldots, a_{q+1}) \in \mathrm{W}(q+1)$, where $a_{q+1} = a_1$. Thus $\mathbf{a}_{q+1}$ corresponds to *closed walk* on G of length q.

Proposition 8.3 *Let $G = (\langle n \rangle, E)$ be a digraph which contains a cycle. Assume that $A = A(G)$ is its representation matrix. Let $\mathrm{W}(q) \supset \mathrm{W}_{\mathrm{per}}(q)$ be the set of all $\mathbf{a}_q$ G-allowable words and the projection of all q-periodic words in $\langle n \rangle^{\mathbb{N}}(G)$ on the first q coordinates. Then $l_q := \mathrm{W}(q) \geq l_{q,\mathrm{per}} := \mathrm{W}_{\mathrm{per}}(q)$*

$$l_{q,\mathrm{per}} = \operatorname{tr} A^q, \ q \in \mathbb{N}, \quad \textit{and} \ \limsup_{q \to \infty} \frac{\log l_{q,\mathrm{per}}}{q} = h(G) = \log \rho(G). \tag{8.7}$$

Proof. Recall that if $A^q = (a_{ij}^{(q)})_{i,j=1}^n$ then $a_{ij}^{(q)}$ is the number of walks from i to j in m steps. (Lemma 4.1.) Hence $\operatorname{tr} A^q = \sum_{i=1}^n a_{ii}^{(q)}$ is the number of closed walks on G of length q, i.e. $\operatorname{tr} A^q = l_{q,\mathrm{per}}$. The second part of (8.7) follows from the second part of (7.2). □

Note that in the second part of (8.7) one can not in general replace lim sup by lim. Indeed, assume that G is strongly connected and $2 \leq p$-periodic. If a_{q+1} is a closed walk on G, of length q, then p divides q. Thus if p does not divide q then $\operatorname{tr} A^q = 0$. This shows that in this case one can not replace lim sup by lim in (8.7). In the case G is strongly connected and aperiodic we can replace lim sup by lim in (8.7). (Use Part 1 of Theorem 7.1.)

8.3 Entropy of Markov Chains

For $x \in \mathbb{R}_+$ let $f(x) = -x \log x$. Then $f(0) = f(1) = 0$, $f(x) > 0$ for $x \in (0, 1)$ and $f(x) < 0$ for $x > 1$. Note that for $x > 0$ $f'' = -\frac{1}{x} < 0$. Hence $f(x)$ is a concave continuous function on $\mathbb{R}_+$. For a distribution $\pi = (\pi_1, \ldots, \pi_n) \in \Pi_n$ the quantity $H(\pi) := -\sum_{i=1}^n \pi_i \log \pi_i$ is called the *entropy* of π. The function H was introduced by Boltzmann, (1844-1906), in Boltzmann H theorem, in his work in statistical mechanics. (Not to be confused with the entropy concept in thermodynamics, introduced by Rudolf Clausius in 1850 and used by Boltzmann.)

In probability $H(\pi)$ measures the uncertainty of the outcome of random random variable X such that $\Pr(X = i) = \pi_i, i = 1, \ldots, n$. It is not difficult to show that $H(\pi) \leq H(\frac{1}{n}\mathbf{1}) = \log n$ and equality holds iff π is the uniform distribution $\frac{1}{n}$. If $\pi = \mathbf{e}_i = (\delta_{i1}, \ldots, \delta_{in})$ then $H(\mathbf{e}_i) = 0$. In this case we know with probability 1 that $X = i$.

$H(\pi)$ measures the capacity of the the channel which transmits the alphabet $\langle n \rangle$ such that the frequency of each letter i is π_i for $i = 1, \ldots, n$. Indeed, assume that we consider all words $\mathbf{a}_m = (a_1, \ldots, a_m) \in \langle n \rangle^m$ such that each letter i appears $m_i \in \mathbb{N}$ times with $m_1 + \ldots + m_n = m$. Then the total number of such words is $t(\mathbf{m}) = \frac{m!}{m_1! \ldots m_n!}$, where $\mathbf{m} = (m_1, \ldots, m_n) \in \mathbb{Z}_+^n$. Let $p_i := \frac{m_i}{m}, i = 1, \ldots, n$ and $\mathbf{p} := (p_1, \ldots, p_n) \in \Pi_n$. Recall Stirling's formula:

$$k! \approx \sqrt{2\pi k} \, \big(\frac{k}{e}\big)^k \quad \texttt{as} \ 1 << k \in \mathbb{N}.$$

Using Stirling's formula, a straightforward calculation shows that

$$\lim_{m\to\infty,\frac{m_i}{m}\to\pi_i,i\in\langle n\rangle}\frac{\log t(\mathbf{m})}{m}=H(\pi). \tag{8.8}$$

Indeed

$$\log t(\mathbf{m})=\log m!-\sum_{i=1}^{n}\log m_i!\approx m\log m-\sum_{i=1}^{n}m_i\log m_i+O(\log m)=$$

$$\sum_{i=1}^{m}-m_i\log\frac{m_i}{m}+O(\log m)=-m\sum_{i=1}^{n}p_i\log p_i+O(\log m).$$

Divide the above formula by m and let $m\to\infty$ to obtain (8.8).

Consider next the Markov chain given by a stochastic matrix $P=(p_{ij})_{i,j=1}^{n}$. Assume that $\mu^{(1)}=(\mu_1^{(1)},\dots,\mu_n^{(1)})\in\Pi$ is a given distribution. Let $X_m\in\langle n\rangle$ be the position of a particle doing the random walk on $G=G(P)=(\langle n\rangle,E)$. We can view this random walk as a walk on the complete directed graph $KD_n=(\langle n\rangle,E)$ such that

$$\Pr(X_m=i_m|X_{m-1}=i_{m-1},X_{m-2}=i_{m-2},\dots,X_1=i_1)=p_{i_{m-1}i_m}.$$

Assume $\Pr(X_1=i)=\mu_i^{(1)}$ for $i=1,\dots,n$. Thus

$$\Pr(X_1=i_1,X_2=i_2,\dots,X_{m-1}=i_{m-1},X_m=i_m)=p_{i_{m-1}i_m}=\mu_{i_1}^{(1)}p_{i_1i_2}\dots p_{i_{m-1}i_m}. \tag{8.9}$$

This is equivalent to saying that the probability of the walk $(i_1,i_2,\dots,i_{m-1},i_m)$ is $\mu_{i_1}^{(1)}\prod_{j=1}^{m-1}p_{i_ji_{j+1}}$.

We now consider the channel with the following property: Given that the letter a_{m-1} was transmitted at the time $m-1$ the probability of the next transmitted letter a_m is given by $p_{a_{m-1}a_m}$. Assume that the probability that the first letter ia a_1 is $\mu_{a_1}^{(1)}$. Then the probability of the word $\mathbf{a}_m=(i_1,\dots,i_m)$ is given by the right-hand side of (8.9). Then the *nomalized* entropy of all words of length m is

$$-\frac{1}{m}\sum_{i_1,\dots,i_m\in\langle n\rangle}\mu_{i_1}^{(1)}\prod_{j=1}^{m-1}p_{i_ji_{j+1}}\log(\mu_{i_1}^{(1)}\prod_{k=1}^{m-1}p_{i_ki_{k+1}})=$$

$$-\frac{1}{m}\sum_{i_1,\dots,i_m\in\langle n\rangle}\mu_{i_1}^{(1)}\prod_{j=1}^{m-1}p_{i_ji_{j+1}}(\log\mu_{i_1}^{(1)}+\sum_{k=1}^{m-1}\log p_{i_ki_{k+1}})=$$

$$-\frac{1}{m}\sum_{i_1,\dots,i_m\in\langle n\rangle}\mu_{i_1}^{(1)}\prod_{j=1}^{m-1}p_{i_ji_{j+1}}\log\mu_{i_1}^{(1)}+$$

$$-\frac{1}{m}\sum_{k=1}^{m-1}(\sum_{i_1,i_{k-1},i_{k+1},\dots,i_m\in\langle n\rangle}\mu_{i_1}^{(1)}\prod_{j=1}^{m-1}p_{i_ji_{j+1}})\log p_{i_ki_{k+1}}=$$

$$-\frac{1}{m}\sum_{i=1}^{n}\mu_i^{(1)}\log\mu_i^{(1)}-\sum_{i,j=1}^{n}\frac{1}{m}\sum_{k=1}^{m-1}(\mu^{(1)}P^{k-1})_ip_{ij}\log p_{ij}.$$

The last equality is deduced as follows. In the expression that involves $\log\mu_{i_1}^{(1)}$ sum on the indices $i_m,i_{m-1},\dots,i_2$ and take into the account that P is stochastic, i.e. $\sum_{i_{j+1}=1}^{n}p_{i_ji_{j+1}}=1$ for $j=m-1,\dots,1$. This gives the first expression in the last equality. To deduce the second part of the equality fix $k\in[1,n]$. Then sum on the indices $i_1,\dots,i_{k-1}$ and use the definition of the product of matrices to obtain the expression $(\mu^{(1)}P^{k-1})_{i_k}$. Then sum on the indices $i_m,\dots,i_{k+2}$ to obtain the expression $(\mu^{(1)}P^{k-1})_{i_k}p_{i_ki_{k+1}}\log p_{i_ki_{k+1}}$ and use the stochasticity of P. Now sum on i_k,i_{k+1} to obtain the second part of the last equality.

Now let $m \to \infty$. Use Corollary 5.3 to deduce that

$$\lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m-1} \mu^{(1)} P^{k-1} = \mu \in \Pi_n, \quad \mu = P\mu. \tag{8.10}$$

Note that any stationary distribution μ of P can be obtained this way, e.g. assume that $\mu^{(1)} = \mu$. Hence the *entropy* of the given Markov chain is given by

$$h(P, \mu) := -\sum_{i,j=1}^{n} \mu_i p_{ij} \log p_{ij}, \tag{8.11}$$

If P has a unique stationary distribution, e.g. P is irreducible, then $h(P) := h(P, \mu)$ is uniquely defined.

Theorem 8.4 *Let $G = (\langle n \rangle, E)$ be a digraph such that the out degree of every vertex in G is positive. Let $A = (a_{ij})_{i,j=1}^n \in \{0,1\}^{n\times n}$ be the representation matrix of G, and denote by $\rho(A)$ the spectral radius of A. Let $P \in [0,1]^{n\times n}$ be a stochastic matrix such that $G(P)$ is a subgraph of G. Let $\mu \in \Pi_n$ be a stationary distribution of P: $\mu P = \mu$. Then $h(P, \mu) \le h(G) = \log \rho(A)$. Furthermore, there exists a stochastic matrix P, ($G(P) \subset G$), and a stationary distribution μ of P such that $h(P, \mu) = h(G)$. If G is strongly connected then P is a unique irreducible stochastic matrix given as follows:*

$$P = (p_{ij})_{i,j=1}^n,\ p_{ij} = \frac{a_{ij}u_j}{\rho(A)u_i},\ i,j = 1,\dots,n,\ A\mathbf{u} = \rho(A)\mathbf{u},\ \mathbf{u} = (u_1,\dots,u_n)^\top > 0. \tag{8.12}$$

The stationary distribution $\mu = (\mu_1, \dots, \mu_n)$ is unique and given by $\mu_i = u_i v_i$ for $i = 1, \dots, n$, where

$$A^\top \mathbf{v} = \rho(A)\mathbf{v}, \quad \mathbf{v} = (v_1,\dots,v_n)^\top > 0, \quad \mathbf{v}^\top \mathbf{u} = 1 \tag{8.13}$$

Proof. The inequality $h(P, \mu) \le h(G)$ follows from the fact that Markov chain is constrained by the Markov condition, hence its capacity can not exceed the Shannon capacity of the channel given by G. Since $\rho(G) = \rho(A)$ is the maximal spectral radius of out of all spectral radii of the strongly connected components of G, it is enough to show that we can find a Markov chain with $h(P, \mu) = h(G)$ in the case where G is strongly connected.

We now show that we have equality for P of the form (8.12). Since G is strongly connected, it contains a cycle. Hence $\rho(A) \ge 1$. As A irreducible the right and the left eigenvectors $\mathbf{u}, \mathbf{v}$ of A corresponding to $\rho(A)$ are unique positive vectors up to a multiplication by a positive scalar. The assumption that $A\mathbf{u} = \rho(A)\mathbf{u}$ implies that P is stochastic, and $G(P) = G$. The equalities (8.13) yield that μ is the unique stationary distribution corresponding to P. We now show that $h(P, \mu) = \log \rho(A)$. Indeed

$$h(P,\mu) = -\sum_{i,j=1}^{n} u_i v_i \frac{a_{ij}u_j}{\rho(A)u_i} \log \frac{a_{ij}u_j}{\rho(A)u_i} = -\sum_{i,j=1}^{n} v_i \frac{a_{ij}u_j}{\rho(A)} \big(\log a_{ij} + \log u_j - \log \rho(A) - \log u_i\big).$$

Since $a_{ij} \in \{0,1\}$ it follows that $a_{ij} \log a_{ij} = 0$ hence all the expressions that contain $a_{ij} \log a_{ij}$ vanish. Observe next

$$\log \rho(A) \sum_{i=1}^{n} \sum_{j=1}^{n} v_i \frac{a_{ij}u_j}{\rho(A)} = \log \rho(A) \sum_{i=1}^{n} v_i u_i = \log \rho(A).$$

Now

$$\sum_{i,j=1}^{n} v_i \frac{a_{ij}u_j}{\rho(A)} \log u_i = \sum_{i=1}^{n} (v_i \sum_{j=1}^{n} \frac{a_{ij}u_j}{\rho(A)}) \log u_i = \sum_{i=1}^{n} v_i u_i \log u_i.$$

Similarly $\sum_{i,j=1}^{n} v_i \frac{a_{ij}u_j}{\rho(A)} \log u_j = \sum_{j=1}^{n} v_j u_j \log u_j$. Hence $h(P, \mu) = \log \rho(A)$.

To show that this is the only case when the equality occurs is more complicated. I know that this result follows from more general result called Parry's theorem [18]. □

8.4 Pressure

This section is an adaptation of some of the notions and the results given in [7]. View a word $\mathbf{a}_m = (a_1, \ldots, a_m) \in \langle n \rangle^m$ as a molecule of length m arranged linearly on the lattice $\mathbb{N} \subset \mathbb{R}$, where the place i is occupies by an atom of a kind $a_i \in \langle n \rangle$ for $i = 1, \ldots, m$. (Thus we have n kinds of different atoms.) We call $\mathbf{a}_m$ an m-configuration. The set of all *unrestricted* m-configurations is $\langle n \rangle^m$. Assume that a potential of each atom of type i is e^{u_i} for some fixed $u_i \in \mathbb{R}$ for $i = 1, \ldots, n$. Then the potential of each m-configuration $\mathbf{a}_m$ is $\prod_{i=1}^n e^{u_{a_i}}$. Let $\mathbf{u} := (u_1, \ldots, u_n)^\top \in \mathbb{R}^n$. Given a digraph $G = (\langle n \rangle, E)$, which has a cycle, let $\mathrm{W}(m)$ be the set of all G-allowable words of length m , i.e. all allowable walks on G of length m. Then the *grand partition function*, (the notion from statistical mechanics), is the sum of all potentials of all G-allowable m-configurations:

$$Z(m, \mathbf{u}, G) := \sum_{\mathbf{a}_m = (a_1, \ldots, a_m) \in \mathrm{W}(m)} \prod_{i=1}^m e^{u_{a_i}}. \tag{8.14}$$

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called nondecreasing if for any $\mathbf{u} = (u_1, \ldots, u_n)^\top \leq \mathbf{v} = (v_1, \ldots, v_n)$, $(u_i \leq v_i, i = 1, \ldots, n)$, $f(\mathbf{u}) \leq f(\mathbf{v})$. Clearly $\prod_{i=1}^m e^{u_{a_i}}$ is a nondecreasing function on $\mathbb{R}^n$ for any $(a_1, \ldots, a_m) \in \langle n \rangle^m$. Hence $Z(m, \mathbf{u}, G)$ and $\log Z(m, \mathbf{u}, G)$ are nondecreasing functions. Let $C \subset \mathbb{R}^n$ be a convex set. Then $f : C \to \mathbb{R}$ is called a *convex* function if $f(t\mathbf{u} + (1-t)\mathbf{v}) \leq tf(\mathbf{u}) + (1-t)f(\mathbf{v})$ for any $\mathbf{u}, \mathbf{v} \in C$ and $t \in [0, 1]$. It is known that $\log Z(m, \mathbf{u}, G)$ is a convex function on $\mathbb{R}^n$ [13].

Theorem 8.5 *Let $G = (\langle n \rangle, E)$ be a directed graph with a cycle. Then the sequence $\log Z(m, \mathbf{u}, G), m = 1, \ldots,$ is a sequence of nondecreasing, convex functions, which is subadditive for each fixed $\mathbf{u} \in \mathbb{R}^n$. Then the pressure $P(\mathbf{u}, G)$ is the defined as the limit $\lim_{m \to \infty} \frac{1}{m} \log Z(m, \mathbf{u}, G)$. $P(\mathbf{u}, G)$ is a nondecreasing, convex, Lipschitz function on $\mathbb{R}^n$:*

$$|P((u_1, \ldots, u_n), G) - P((v_1, \ldots, v_n), G)| \leq \max_{i \in [1,n]} |u_i - v_i|. \tag{8.15}$$

Denote by $A = (a_{ij})_{i,j=1}^n \in \{0, 1\}^{n \times n}$ the representation matrix of G. Let $A(\mathbf{u}) := (a_{ij} e^{\frac{u_i + u_j}{2}})_{i,j=1}^n \in \mathbb{R}_+^{n \times n}$. Denote by $\rho(\mathbf{u})$ the spectral radius of $A(\mathbf{u})$. Then $P(\mathbf{u}, G) = \log \rho(\mathbf{u})$. If G is strongly connected then $\log \rho(\mathbf{u})$ is a smooth function of $\mathbf{u}$.

Proof. The sequence $\log Z(m, \mathbf{u}, G), m = 1, \ldots,$ is subadditive for the same reason $\log l_m = \log Z(m, \mathbf{0}, G)$ is subadditive. Hence the sequence $\frac{1}{m} \log Z(m, \mathbf{u}, G)$ converges to $P(\mathbf{u}, G)$ for each $\mathbf{u} \in \mathbb{R}^n$. Since each $\frac{1}{m} \log Z(m, \mathbf{u}, G)$ is nondecreasing and convex on $\mathbb{R}^n$ it follows that $P(\mathbf{u}, G)$ is nondecreasing and convex. Fix $\mathbf{u} = (u_1, \ldots, u_n)^\top, \mathbf{v} = (v_1, \ldots, v_n)^\top \in \mathbb{R}^n$. Let $t = \max_{i \in [1,n]} |u_i - v_i|$. Then $v_i - t \leq u_i \leq v_i + t$ for all $i = 1, \ldots, n$. Hence

$$Z(m, \mathbf{v}, G) e^{-mt} \leq Z(m, \mathbf{u}, G) \leq Z(m, \mathbf{v}, G) e^{mt} \Rightarrow |\frac{1}{m} \log Z(m, \mathbf{u}, G) - \frac{1}{m} \log Z(m, \mathbf{v}, G)| \leq t.$$

Let $m \to \infty$ and deduce (8.15). We now compare $\mathbf{1}^\top A(\mathbf{u})^{m-1} \mathbf{1}$ with $Z(m, \mathbf{u}, G)$. One term in $\mathbf{1}^\top A(\mathbf{u})^{m-1} \mathbf{1}$ is of the form $e^{\frac{u_{i_1}}{2}} a_{i_1 i_2} e^{\frac{u_{i_2}}{2}} e^{\frac{u_{i_2}}{2}} a_{i_2 i_3} e^{\frac{u_{i_3}}{2}} \ldots e^{\frac{u_{i_{m-1}}}{2}} a_{i_{m-1} i_m} e^{\frac{u_{i_m}}{2}}$. If $(i_1, i_2, \ldots, i_m) \notin \mathrm{W}(m)$ then this product is equal to zero. If $(i_1, i_2, \ldots, i_m) \in \mathrm{W}(m)$ then this product is equal to $e^{-(\frac{u_{i_1}}{2} + \frac{u_{i_m}}{2})} \prod_{j=1}^m e^{u_{i_j}}$. Let $t = \max_{i \in [1,n]} |u_i|$. Then $Z(m, \mathbf{u}, G) e^{-t} \leq \mathbf{1}^\top A^{m-1} \mathbf{1} \leq Z(m, \mathbf{u}, G) e^t$. Take the logarithm in all the terms of this inequality, divide by, let $m \to \infty$ and use (7.2) to deduce the equality $P(\mathbf{u}, G) = \log \rho(A(\mathbf{u}))$.

Assume finally that G is strongly connected. Then A is irreducible, hence $A(\mathbf{u})$ is irreducible. So $\rho(\mathbf{u}) > 0$ is a simple root of the characteristic equation of $\det(zI - A(\mathbf{u}))$. Note that coefficient of this characteristic polynomial are analytic functions of $\mathbf{u}$. Hence the implicit function theorem yields that $\rho(\mathbf{u})$ and hence $\log \rho(\mathbf{u})$ are smooth functions of $\mathbf{u}$. □

For $m \in \mathbb{N}$ denote

$$\mathcal{C}_n(m) := \{\mathbf{c} = (c_1, \ldots, c_n) \in \mathbb{Z}_+^n : c_1 + \cdots + c_n = m\}.$$

or any G-allowable word $\mathbf{a}_m = (a_1, \ldots, a_m) \in \mathrm{W}(m)$, let $\mathbf{c}(\mathbf{a}_m) = (c_1, \ldots, c_n) \in \mathcal{C}_n(m)$ the frequency vector of the letter distributions in $\mathbf{a}_m$. That is c_i is the number the letter i appears in $\mathbf{a}_m$. For any $\mathbf{c} \in \mathcal{C}_n(m)$ set

$$\mathrm{W}(\mathbf{m}, \mathbf{c}) := \{\mathbf{a}_m \in \mathrm{W}(\mathbf{m}) : \mathbf{c}(\mathbf{a}_m) = \mathbf{c}\}, \quad \text{for all } \mathbf{c} \in \mathcal{C}_n(m).$$

This is the set of G-allowable words of $\mathbf{a}_m \in \mathrm{W}(m)$ with color frequency vector $\mathbf{c}$.

Definition 8.6 *$\mathbf{p} \in \Pi_n$ is called a density point of $\langle n \rangle^{\mathbb{N}}(G)$, when there exist an increasing sequences of natural integers $m_q \in \mathbb{N}$ and color frequency vectors $\mathbf{c}_q \in \mathcal{C}_n(m_q)$ such that*

$$m_q \to \infty, \ \ \mathrm{W}(m_q, \mathbf{c}_q) \neq \emptyset \ \ \forall q \in \mathbb{N}, \quad \textit{and} \ \lim_{q\to\infty} \frac{\mathbf{c}_q}{m_q} = \mathbf{p}. \tag{8.16}$$

We denote by Π_G the set of all density points of $\langle n \rangle^{\mathbb{N}}(G)$. For $\mathbf{p} \in \Pi_G$ we let

$$h_G^*(\mathbf{p}) := \sup_{m_q, \mathbf{c}_q} \limsup_{q\to\infty} \frac{\log \#\mathrm{W}(m_q, \mathbf{c}_q)}{m_q} \geq 0, \tag{8.17}$$

where the supremum is taken over all sequences satisfying (8.16). One can think of $h_G^*(\mathbf{p})$ as the entropy for the color density $\mathbf{p}$.

It is straightforward to show (using a variant of the Cantor diagonal argument) that Π_G is a closed set. Furthermore, h_G^* is *upper semi-continuous* on Π_G.

Theorem 8.7 *Let $G = (\langle n \rangle, E)$ be a strongly connected graph. Let $P(\cdot, G) : \mathbb{R}^n \to \mathbb{R}$ be the pressure function associate with G. Then $\nabla P(\mathbf{u}) := (\frac{\partial P}{\partial u_1}(\mathbf{u}), \ldots, \frac{\partial P}{\partial u_n}(\mathbf{u})) \ (\in \Pi_n)$ is a distribution for each $\mathbf{u} \in \mathbb{R}^n$. Furthermore, $\nabla P(\mathbf{u}) \in \Pi_G$ and*

$$h_G^*(\nabla P(\mathbf{u})) = P(\mathbf{u}) - \nabla P(\mathbf{u})\mathbf{u}, \tag{8.18}$$

$$P(\mathbf{u}) = \max_{\mathbf{p}\in\Pi_G} (\mathbf{p}\mathbf{u} + h_G^*(\mathbf{p})), \tag{8.19}$$

for each $\mathbf{u} \in \mathbb{R}^n$.

Proof. We first show that $\nabla P(\mathbf{u}) \in \Pi_n$. Since $P(\mathbf{u}, G)$ is nondecreasing it follows that $\nabla P(\mathbf{u}) \geq \mathbf{0}$. Let $f(t, \mathbf{u}) = P(\mathbf{u} + t\mathbf{1})$ for any $t \in \mathbb{R}$. From the definition of $A(\mathbf{u})$ it follows that $A(\mathbf{u} + t\mathbf{1}) = e^t A(\mathbf{u})$. Hence

$$f(t, \mathbf{u}) = P(\mathbf{u} + t\mathbf{1}) = \log \rho(A(\mathbf{u} + t\mathbf{1})) = t + \log \rho(\mathbf{u}) = t + P(\mathbf{u}).$$

Fix $\mathbf{u}$ and take the derivative with respect to t. The chain rule implies $1 = \frac{df(t,\mathbf{u})}{dt} = \nabla P(\mathbf{u})\mathbf{1}$, which implies that $\nabla P(\mathbf{u}) \in \Pi_n$.

We now show the inequality

$$P(\mathbf{u}) \geq \mathbf{p}\mathbf{u} + h_G^*(\mathbf{p}) \quad \texttt{for any } \mathbf{p} \in \Pi_G \texttt{ and } \mathbf{u} \in \mathbb{R}^n. \tag{8.20}$$

Fix $\mathbf{p} \in \Pi_G$ and let $m_q, \mathbf{c}_q$, $q \in \mathbb{N}$, be sequences satisfying (8.16). We have $Z(m_q, \mathbf{u}, G) \geq \#\mathrm{W}(m_q, \mathbf{c}_q) e^{\mathbf{c}_q \mathbf{u}}$, since the right-hand side is just a partial sum of the sum represented by left-hand side. Take logarithms, divide by m_q, take $\limsup_{q\to\infty}$ and use the definition of $P(\mathbf{u})$ and the limit in (8.16) to deduce $P(\mathbf{u}) \geq \mathbf{p}\mathbf{u} + \limsup_{q\to\infty} \frac{\log \#\mathrm{W}(m_q, \mathbf{c}_q)}{m_q}$. Now take the supremum over all sequences $m_q, \mathbf{c}_q$ satisfying (8.16) and use (8.17) to obtain (8.20). Hence

$$P(\mathbf{u}, G) \geq \sup_{\mathbf{p}\in\Pi_G} \mathbf{p}\mathbf{u} + h_{\mathbf{\Gamma}}^*(\mathbf{p}). \tag{8.21}$$

We now show that for each $\mathbf{u} \in \mathbb{R}^n$ there exists $\mathbf{p}(\mathbf{u}) \in \Pi_G$ such that

$$P(\mathbf{u}, G) = \mathbf{p}(\mathbf{u})\mathbf{u} + h^*_G(\mathbf{p}(\mathbf{u})). \tag{8.22}$$

Observe first that

$$\#\mathcal{C}_n(m) = \binom{m+n-1}{n-1} = O(m^{n-1}), \quad m \to \infty.$$

Then for each $m \in \mathbb{N}^d$

$$Z(m, \mathbf{u}, G) = O(m^{n-1}) \max_{\mathbf{c} \in \Pi_n(m)} \#\mathrm{W}(m, \mathbf{c}) e^{\mathbf{c}^\top \mathbf{u}}.$$

Let

$$\mathbf{c}(m, \mathbf{u}) := \arg \max_{\mathbf{c} \in \Pi_n(m)} \#\mathrm{W}(m, \mathbf{c}) e^{\mathbf{c}\mathbf{u}}. \tag{8.23}$$

Then

$$Z(m, \mathbf{u}, G) = O(m^{n-1}) \#\mathrm{W}(m, \mathbf{c}(m, \mathbf{u})) e^{\mathbf{c}(m,\mathbf{u})\mathbf{u}}. \tag{8.24}$$

Choose a sequence m_q such that $\frac{\mathbf{c}(m_q, \mathbf{u})}{m_q}$ converges to some $\mathbf{p}(\mathbf{u})$. We have $\mathbf{p}(\mathbf{u}) \in \Pi_G$ by Definition 8.16. Apply (8.24) to m_q, and use the definition of $P(\mathbf{u}, G)$, $h^*_G(\mathbf{p}(\mathbf{u}))$ to deduce

$$P(\mathbf{u}, G) \le \mathbf{p}(\mathbf{u})\mathbf{u} + \limsup_{l \to \infty} \frac{\log \#\mathrm{W}(m_q, \mathbf{c}(m_q))}{\mathbf{m}_q} \le \mathbf{p}(\mathbf{u})\mathbf{u} + h^*_G(\mathbf{p}(\mathbf{u})).$$

Combine (8.21) with the above inequality to deduce (8.22) and(8.19).

It is left to show that $\nabla P(\mathbf{u}) = \mathbf{p}(\mathbf{u})$. Let $\mathbf{v} \in \mathbb{R}^n$ and $t \in \mathbb{R}$. Since $\mathbf{p}(\mathbf{u}) \in \Pi_G$ the inequality (8.21) combined with the quality (8.22) yields that

$$P(\mathbf{u}+t\mathbf{v}, G) \ge \mathbf{p}(\mathbf{u})(\mathbf{u}+t\mathbf{v}) + h^*_G(\mathbf{p}(\mathbf{u})) = t\mathbf{p}(\mathbf{u})\mathbf{v} + P(\mathbf{u}, G) \Rightarrow P(\mathbf{u}+t\mathbf{v}, G) - P(\mathbf{u}, G) \ge t\mathbf{p}(\mathbf{u})\mathbf{v}.$$

Assume that $t > 0$. Divide by t and let $t \searrow 0$ to deduce that $\nabla P(\mathbf{u})\mathbf{v} \ge \mathbf{p}(\mathbf{u})\mathbf{v}$. Assume that $t < 0$. Divide by t and let $t \nearrow 0$ to deduce that $\nabla P(\mathbf{u})\mathbf{v} \le \mathbf{p}(\mathbf{u})\mathbf{v}$. Hence $\nabla P(\mathbf{u})\mathbf{v} = \mathbf{p}(u)\mathbf{v}$for all $\mathbf{v} \in \mathbb{R}^n$ which implies that $\nabla P(\mathbf{u}) = \mathbf{p}(\mathbf{u})$. Use (8.22) to deduce (8.18). □

Proposition 8.8 *Let $G = (\langle n \rangle, E)$ be a strongly connected graph. For each $\mathbf{u} \in \mathbb{R}^n$ let $A(\mathbf{u}) \in \mathbb{R}^{n \times n}_+$ be defined as in Theorem 8.5. Assume that $\mathbf{x}(\mathbf{u}) = (x_1(\mathbf{u}), \dots, x_n(\mathbf{u}))^\top, \mathbf{y}(\mathbf{u}) = (y_1(\mathbf{u}), \dots, y_n(\mathbf{u}))^\top \in \mathbb{R}^n_+$ be positive left and right eigenvectors of $A(\mathbf{u})$: $A(\mathbf{u})\mathbf{x}(\mathbf{u}) = \rho(\mathbf{u})\mathbf{x}(\mathbf{u}), A(\mathbf{u})^\top \mathbf{y}(\mathbf{u}) = \rho(\mathbf{u})\mathbf{y}(\mathbf{u})$ normalized by the condition $\mathbf{y}(\mathbf{u})^\top \mathbf{x}(\mathbf{u}) = 1$. Then $\nabla P(\mathbf{u}) = (y_1(\mathbf{u})x_1(\mathbf{u}), \dots, y_n(\mathbf{u})x_n(\mathbf{u}))$ for each $\mathbf{u} \in \mathbb{R}^n$.*

Proof. Since $\rho(\mathbf{u}) > 0$ is a simple root of $\det(zI - A(\mathbf{u}))$ it follows that one can choose $\mathbf{x}(\mathbf{u}), \mathbf{y}(\mathbf{u})$ to be analytic on $\mathbb{R}^n$ in $\mathbf{u}$. (For example first choose $\mathbf{x}(\mathbf{u}), \tilde{\mathbf{y}}(\mathbf{u}) \in \mathbb{R}^n_+$ to be the unique left and right eigenvectors of $A(\mathbf{u})$ of length 1. Then let $\mathbf{y}(\mathbf{u}) = \frac{1}{\tilde{\mathbf{y}}(\mathbf{u})^\top \mathbf{x}(\mathbf{u})} \mathbf{y}(\mathbf{u})$. Let ∂_i be the partial derivative with respect to u_i. Then

$$\mathbf{y}(\mathbf{u})^\top \mathbf{x}(\mathbf{u}) = 1 \ \texttt{for all} \ \mathbf{u} \in \mathbb{R}^n \Rightarrow \partial_i \mathbf{y}(\mathbf{u})^\top \mathbf{x}(\mathbf{u}) + \mathbf{y}(\mathbf{u})^\top \partial_i \mathbf{x}(\mathbf{u}) = 0, \ \texttt{for} \ i = 1, \dots, n.$$

Observe next that $\mathbf{y}(\mathbf{u})^\top A(\mathbf{u})\mathbf{x}(\mathbf{u}) = \rho(\mathbf{u})$. Taking the partial derivative with respect to u_i we get

$$\partial_i \rho(\mathbf{u}) = \partial_i \mathbf{y}(\mathbf{u})^\top A(\mathbf{u})\mathbf{x}(\mathbf{u}) + \mathbf{y}(\mathbf{u})^\top A(\mathbf{u}) \partial_i \mathbf{x}(\mathbf{u}) + \mathbf{y}(\mathbf{u})^\top \partial_i A(\mathbf{u})\mathbf{x}(\mathbf{u}).$$
$$\rho(\mathbf{u})(\partial_i \mathbf{y}(\mathbf{u})^\top \mathbf{x}(\mathbf{u}) + \mathbf{y}(\mathbf{u})^\top \partial_i \mathbf{x}(\mathbf{u})) + \rho(\mathbf{u}) y_i(\mathbf{u}) x_i(\mathbf{u}) = \rho(\mathbf{u}) y_i(\mathbf{u}) x_i(\mathbf{u}).$$

Recalling the equality $P(\mathbf{u}, G) = \log \rho(\mathbf{u})$ we deduce that $\nabla P(\mathbf{u}) = (y_1(\mathbf{u})x_1(\mathbf{u}), \dots, y_n(\mathbf{u})x_n(\mathbf{u}))$. □

Since $h(G) = P(\mathbf{0}, G)$ the characterization (8.19) yields:

$$h(G) = \max_{\mathbf{p} \in \Pi_G} h_G^*(\mathbf{p}). \tag{8.25}$$

Let $A \in [0,1]^{n \times n}$ be an irreducible matrix. Let $\mathbf{x} = (x_1, \ldots, x_n)^\top, \mathbf{y} = (y_1, \ldots, y_n)^\top \in \mathbb{R}_+^n$ be the left and the right eigenvectors of A corresponding to $\rho(A)$, and normalized by the condition $\mathbf{y}^\top \mathbf{x} = 1$. Then the distribution $(y_1 x_1, \ldots, y_n x_n) \in \Pi_n$ appears naturally in other problems as well. For example the Friedland-Karlin characterization of $\log \rho(A)$ is [8]:

$$\log \rho(A) = \min_{\mathbf{z}=(z_1,\ldots,z_n)^\top > 0} \sum_{i=1}^n y_i x_i \log \frac{(A\mathbf{z})_i}{z_i}. \tag{8.26}$$

We now apply Theorem 8.7 to the hard core model on $\langle 2 \rangle^{\mathbb{N}}(G)$, where $A(G) = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. (We identified 0 with 2.) Let $\mathbf{u} = (s,t)^\top$. Then $\nabla P(\mathbf{u}, G) = (p_1(\mathbf{u}), p_2(\mathbf{u})) \in \Pi_2$ it follows that $p_2(\mathbf{u}) = 1 - p_1(\mathbf{u})$. It is enough to consider $\mathbf{u} = (s,0)$ and $p_1(s) = \frac{dP((s,0)^\top, G)}{ds}$. So $p := p_1(s)$ is the density of 1 in all the configurations of infinite strings of $0, 1$, where no two 1 are adjacent. Clearly $A(\mathbf{u}) = \begin{pmatrix} 0 & e^{\frac{s}{2}} \\ e^{\frac{s}{2}} & 1 \end{pmatrix}$. Hence

$$\rho(\mathbf{u}) = \frac{1 + \sqrt{1 + 4e^s}}{2}, \quad p_1(s) = \frac{2e^s}{(1 + \sqrt{1 + 4e^s})\sqrt{1 + 4e^s}} = \frac{2}{(e^{-\frac{s}{2}} + \sqrt{e^{-s} + 4})\sqrt{e^{-s} + 4}} = \frac{1}{2}\big(1 - \frac{1}{\sqrt{1 + 4e^s}}\big) \in (0, \frac{1}{2}).$$

Note that $p_1(s)$ is increasing on $\mathbb{R}$, and $p_1(-\infty) = 0, p_1(\infty) = \frac{1}{2}$. As $P(\mathbf{0}) = h(G) = \log \frac{1+\sqrt{5}}{2}$ it follows that the value $p^* := p_1(0) = \frac{2}{(1+\sqrt{5})\sqrt{5}} = .2763932024$ is the density p^* of 1's for which $h(G) = h_G^*(p^*)$. To find the formula for $h_G(p)$ first note that if $p = p_1(s)$ then

$$\sqrt{1 + 4e^s} = \frac{1}{1 - 2p}, \quad s(p) = \log \frac{p(1-p)}{(1-2p)^2}.$$

Then

$$h_G^*(p) = \log \frac{1-p}{1-2p} - p \log \frac{p(1-p)}{(1-2p)^2}, \quad p \in (0, \frac{1}{2}).$$

9 Simulation of the gradient of pressure

In this section we show that the gradient of a pressure is easy to simulate. Let

$$P_m(\mathbf{u}, G) := \frac{\log Z(m, \mathbf{u}, G)}{m}, \quad m \in \mathbb{N}. \tag{9.1}$$

Then

$$\frac{\partial P_m(\mathbf{u}, G)}{\partial u_i} = \frac{1}{m} \sum_{\phi \in \mathrm{W}(m)} c_i(\phi) \frac{e^{\mathbf{c}(\phi)\mathbf{u}}}{Z(m, \mathbf{u}, G)}. \tag{9.2}$$

On $\mathrm{W}(m)$ we introduce that following probability which depends on $\mathbf{u}$:

$$\Pr{}_{\mathbf{u}}(\phi) := \frac{e^{\mathbf{c}(\phi)\mathbf{u}}}{Z(m, \mathbf{u}, G)}, \quad \texttt{for any } \phi \in \mathrm{W}(m). \tag{9.3}$$

Let $X_{i,m} : \mathrm{W}(m) \to \mathbb{Z}_+$ be the random variable that counts the number of atoms, (letters), of type i in the state $\phi \in \mathrm{W}(m)$. That is $X_{i,m} = c_i(\phi)$ for $i = 1, \ldots, n$. Let $\mathbf{X}_m :=$

$(X_{1,m},\dots,X_{n,m}):\mathrm{W}(m)\to\mathbb{Z}_+^m$ be a vector random variable. Then $\mathbf{X}_m(\phi)=\mathbf{c}(\phi)$. Use (9.1) and (9.3) to deduce that

$$\nabla P_m(\mathbf{u},G)=\frac{1}{m}\mathrm{E}_{\mathbf{u}}(\mathbf{X}_m),\quad m\in\mathbb{N}. \tag{9.4}$$

Note that in HW 3 we considered the computation of the expected number of $1's$ in the hard core configurations. This quantity corresponds to $E_{\mathbf{0}}(X_{1,m})$. Second part of the Problem 4 was to show that $\frac{E_{\mathbf{0}}(X_{1,m})}{m}$ converges to a limit. It can be shown that this limit is given by $p_1(0)=\frac{2}{(1+\sqrt{5})\sqrt{5}}$, where $p_1(s)$ is defined in the last part of the previous section. More precisely, if G is strongly connected it can be shown that $\lim_{m\to\infty}\nabla P_m(\mathbf{u},G)=\nabla P(\mathbf{u},G)$.

To find the approximate value of $\nabla P_m(\mathbf{u},G)$ we simulate a random walk on G using a Markov chain as follows. (We assume that G is strongly connected and $m>1$.)
Variant 1: Start the program with $w=e_1=\ldots=e_n=0$. Let $P=(p_{\mathbf{a}_m\mathbf{b}_m})_{\mathbf{a}_m,\mathbf{b}_m\in\mathrm{W}(m)}$ be the following stochastic matrix on the graph $G_m=(\mathrm{W}(m),E_m)$, with vertices indexed by all allowable m - walks $\mathrm{W}(m)$. $(\mathbf{a}_m,\mathbf{b}_m)\in E_m$ if and only if $\mathbf{a}_m=(a_1,\dots,a_m),\mathbf{b}_m=(b_1,\dots,b_m)\in\mathrm{W}(m)$ differ at most in one vertex k, for some $k=1,\dots,n$. Clearly G_m is reversible, i.e. $(\mathbf{a}_m,\mathbf{b}_m)\in E_m\iff(\mathbf{b}_m,\mathbf{a}_m)\in E_m$. Assume that $a_1=b_1,\dots,a_{k-1}=b_{k-1},a_{k+1}=b_{k+1},\dots,a_m=b_m$. Let $r=r(a_{k-1},a_{k+1})$ be the number of all $j\in\langle n\rangle$ such that

$$(a_{k-1},j)\in E(G)\ \texttt{and}\ (j,a_{k+1})\in E(G),\quad \texttt{for all}\ j\in\langle n\rangle. \tag{9.5}$$

If $k=1$ then the above condition reduces to $(j,a_2)\in E(G)$. If $k=m$ the above condition reduces to $(a_{m-1},j)\in E(G)$. Then for $a_k\neq b_k$, (hence $a_i=b_i$ for $i\neq k$), we have that $p_{\mathbf{a}_m\mathbf{b}_m}=\frac{1}{mr(a_{k-1},a_{k+1})}$. $p_{\mathbf{a}_m\mathbf{a}_m}$ is determined by the stochasticity condition. Then P is a symmetric stochastic matrix. If G_m is connected then P has a unique uniform distribution. (We do not in general that G_m is connected, but it is straightforward to show that G_m is connected for the hard core model.)

Our random walk on G_m is given as follows. First generate a walk of length $\mathbf{a}_m$. Let $v=\mathbf{c}(\mathbf{a}_m)\mathbf{u}=c_1(\mathbf{a}_m)u_1+\ldots c_n(\mathbf{a}_m)u_n$, where $c_i(\mathbf{a}_m)$ is the number of i vertices in the walk given by $\mathbf{a}_m$. Let $s=\exp(v)$, $w=w+s$ and $e_i=e_i+c_i(\mathbf{a}_m)s$ for $i=1,\dots,m$.

Then move form $\mathbf{a}_m$ to $\mathbf{b}_m$ as follows. Choose a vertex a_k for some $k=1,\dots,m$ with probability $\frac{1}{m}$. Then choose $b_k=j$, where (9.5) holds, with probability $\frac{1}{r(a_{k-1},a_{k+1})}$. Thus $\mathbf{b}_m=(a_1,\dots,a_{k-1},b_k,a_{k+1},\dots,a_m)$. Then $c_i(\mathbf{b}_m)=c_i(\mathbf{a}_m)-c_i(a_k)+c_i(b_k)$ for $i=1,\dots,n$. (Here $c_i(j)$ is the number of the color i in the color j. So $c_i(j)=\delta_{ij}$.) Now find $s,w,\mathbf{e}=(e_1,\dots,e_n)$ by replacing $\mathbf{a}_m$ by $\mathbf{b}_m$ in the above equalities, and continue.

When you finished the iterations the vector $\frac{1}{w}\mathbf{e}$ is an estimate of $\mathrm{E}(\mathbf{X}_m)$ and $\frac{1}{mw}\mathbf{e}$ is an estimate of $\nabla P_m(\mathbf{u},G)$. (If G_m is connected it is straightforward to show that these are valid estimates.)
Variant 2: Start the program with $w=e_1=\ldots=e_n=0$. Let $P=(p_{ij})_{i,j=1}^n$ where $p_{ij}=\frac{a_{ij}x_j}{\rho(G)x_i}$, where $A(G)=(a_{ij})_{i,j=1}^n\in\{0,1\}^{n\times n}$ is the incidence matrix of G and $\mathbf{x}=(x_1,\dots,x_n)^\top\in\mathbb{R}_+^n$ is the positive eigenvector of $A(G)$ corresponding to $A(G)$: $A(G)\mathbf{x}=\rho(G)\mathbf{x}$. Choose a vertex a_1 at random with probability $\frac{1}{n}$. Then choose a neighbor of a_2 with probability $p_{a_1a_2}$ and so on until one gets a word $\mathbf{a}_m=(a_1,\dots,a_m)\in\mathrm{W}(m)$. Then let $v=\mathbf{c}(\mathbf{a}_m)\mathbf{u}=c_1(\mathbf{a}_m)u_1+\ldots c_n(\mathbf{a}_m)u_n$, where $c_i(\mathbf{a}_m)$ is the number of i vertices in the walk given by $\mathbf{a}_m$. Let $s=\frac{x_{a_1}\exp(v)}{x_{a_m}}$, $w=w+s$ and $e_i=e_i+c_i(\mathbf{a}_m)s$ for $i=1,\dots,m$.

Pick a vertex a_{m+1}, such $(a_m,a_{m+1})\in E(G)$, and define the configuration $\mathbf{b}_m:=(a_2,\dots,a_{m+1})$. Then $c_i(\mathbf{b}_m)=c_i(\mathbf{a}_m)-c_i(a_1)+c_i(a_{m+1})$ for $i=1,\dots,n$. (Here $c_i(j)$ is the number of the color i in the color j. So $c_i(j)=\delta_{ij}$.) Now find $s,w,\mathbf{e}=(e_1,\dots,e_n)$ by replacing $\mathbf{a}_m$ by $\mathbf{b}_m$ in the above equalities, and continue.

When you finished the iterations the vector $\frac{1}{w}\mathbf{e}$ is an estimate of $\mathrm{E}(\mathbf{X}_m)$ and $\frac{1}{mw}\mathbf{e}$ is an estimate of $\nabla P_m(\mathbf{u},G)$.

We now explain why $\frac{1}{w}\mathbf{e}$ is an estimate of $\mathrm{E}(\mathbf{X}_m)$ in this case, i.e. Variant 2. Indeed the probability of generating a word $\mathbf{a}_m = (a_1, \ldots, a_m)$ is $\frac{1}{n}p_{a_1a_2}p_{a_2a_3}\cdots p_{a_{m-1}a_m} = \frac{1}{n}\frac{x_{a_m}}{\rho(G)^{m-1}x_{a_1}}$. This explains why we consider the product $s = \frac{x_{a_1}\exp(v)}{x_{a_m}}$. Hence $\frac{1}{w}\mathbf{e}$ is an estimate of $\mathrm{E}(\mathbf{X}_m)$.

It is straightforward to show that if under the assumption that G is strongly connected you can go from any $\mathbf{a}_m \in \mathrm{W}(m)$ to any other $\mathbf{b}_m \in \mathrm{W}(m)$ using the Markov chain described above in a finite number of steps. The disadvantage Variant 2 is that we need to know the eigenvector $\mathbf{x}$ and $\rho(G)$.

Remark 9.1 *Assume that $G = (\langle n\rangle, E)$ is strongly connected. $A(G) = (a_{ij})_{i,j=1}^n \in \{0,1\}^{n\times n}$ is the incidence matrix of G and $\mathbf{x} = (x_1, \ldots, x_n)^\top, \mathbf{y} = (y_1, \ldots, y_n)^\top \in \mathbb{R}_+^n$ is the positive left and right eigenvectors of $A(G)$ corresponding to $A(G)$: $A(G)\mathbf{x} = \rho(G)\mathbf{x}, \mathbf{y}^\top A(G) = \rho(G)\mathbf{y}^\top$, normalized by the condition $\mathbf{y}^\top\mathbf{x} = 1$. Introduce on $\langle n\rangle^m$, the sets of walks of length m on the complete digraph on n vertices the following probability measure:*

$$\Pr(\mathbf{b}_m = (b_1, \ldots, b_m)) := \frac{1}{\rho(G)^{m-1}} y_{b_1} a_{b_1b_2}\ldots a_{b_{m-1}b_m} x_{b_m}, \quad \texttt{for all } b_1, \ldots, b_m \in \langle n\rangle. \tag{9.6}$$

Then $\Pr(\mathbf{b}) = 0$ if $\mathbf{b} \notin \mathrm{W}(m)$, and $\Pr(\mathbf{b}) = \frac{1}{\rho(G)^{m-1}} y_{b_1}x_{b_m}$ if $\mathbf{b}_m = (b_1, \ldots, b_m) \in \mathrm{W}(m)$. This measure is called Parry measure, and is an example of Gibbs measure.

References

[1] D. Aldous and J.A. Fill:, *Reversible Markov Chains and Random Walks on Graphs*, monograph in preparation, http://stat-www.berkeley.edu/∼aldous/RWG/book.html.

[2] N. Alon and J. Spencer, *The Probabilistic Method*, Wiley, 1992.

[3] R.B. Bapat and T.E.S. Raghavan, *Nonnegative Matrices and Applications*, Cambridge University Press, Cambridge, UK, 1997.

[4] B. Bollobás, *Graph Theory*, Springer-Verlag, 1979.

[5] B. Bollobás, *Modern Graph Theory*, Springer-Verlag, 1998.

[6] P. Erdös and A. Renyi, On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kut. Int. Kozl.*, 5 (1960), 17-61.

[7] S. Friedland and U.N. Peled, The pressure associated with multidimensional SOFT, in preparation.

[8] S. Friedland and S. Karlin, Some inequalities for the spectral radius of nonnegative matrices and applications, *Duke Math. J.* 42 (1975), 459-490.

[9] S. Friedland and H. Schneider, The growth of powers of nonnegative matrix, *SIAM J. Algebraic Discrete Methods* 1 (1980), 185-200.

[10] O. Häggström, *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press, Cambridge, UK, 2002.

[11] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.

[12] M. Jerrum, *Counting, Sampling and Integrating: Algorithms and Complexity*, Springer Verlag, 2003. http://www.dcs.ed.ac.uk/home/mrj/pubs.html

[13] J.F.C. Kingman, A convexity property of positive matrices. *Quart. J. Math. Oxford Ser.* (2) 12 (1961), 283–284.

[14] D.P. Landau and K. Binder, *A Guide to Monte Carlo Simulaitons in Statistical Physics*, Cambridge University Press, Cambridge, UK, 2000.

[15] S.J. Leon, *Linear Algebra*, Prentice Hall, 2002, 6-th. ed.

[16] R. Lyons with Y. Peres, *Probability on Trees and Networks*, http://mypage.iu.edu/∼rdlyons/prbtree/prbtree.html

[17] H. Minc, *Nonnegative Matrices*, Wiley 1988.

[18] W. Parry, Intrinsic Markov Chains, *Trans. Amer. Math. Soc.* 112 (1964), 5-65.

[19] U.G. Rothblum, Algebraic eigenspaces of nonnegative matrices, *Linear Algebra Appl.* 12 (1975), 281-292.

[20] E. Seneta, *Non-Negative Matrices and Markov Chains*, 2nd ed. Springer, New York, 1981

[21] A. Sinclair, *Algorithms for random generation and counting: a Markov chain approach*, Birkhauser, 1993.

[22] J. Spencer, *Ten Lectures on the Probabilistic Method*, SIAM, Philadelphia, 1987.

[23] P. Walters, *An Introduction to Ergodic Theory*, Springer-Verlag, 1982.