*Stochastic Processes and their Applications 2 (1974) 311–336. © North-Holland Publishing Company*

# RANDOM WALKS ON GRAPHS

F. GÖBEL and A.A. JAGERS
*Technische Hogeschool Twente, Enschede, The Netherlands*

Received 19 March 1973
Revised 15 January 1974

**Abstract.** In this paper the following Markov chains are considered: the state space is the set of vertices of a connected graph, and for each vertex the transition is always to an adjacent vertex, such that each of the adjacent vertices has the same probability. Detailed results are given on the expectation of recurrence times, of first-entrance times, and of symmetrized first-entrance times (called commuting times). The problem of characterizing all connected graphs for which the commuting time is constant over all pairs of adjacent vertices is solved almost completely.

| | |
|---|---|
| balanced graph | random walk |
| block (of a graph) | tree-wise join |
| first entrance time | |

## 1. Preliminaries

### *1.1. Intuitive introduction and summary*

Consider the graph of Fig. 1.1. At A we start a random walk along the edges. After one step we are at the vertex B. The next step is either BA or BC. We assign probability $\frac{1}{2}$ to each of the two possibilities. Once we are at C, there are three possibilities for the next step: CB, CD or CE. To each we assign probability $\frac{1}{3}$. We agree to stop as soon as we are at D. The

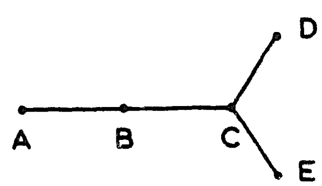


Fig. 1.1. A simple maze.

expected number of steps in this random walk, to be denoted by $\delta_{AD}$, can be easily calculated: $\delta_{AD} = 11$. The analogous quantity $\delta_{DA}$ has the value 13. If $i$ and $j$ are two vertices of the above graph, we may consider $\gamma_{ij}$ defined by

$$\gamma_{ij} = \delta_{ij} + \delta_{ji}.$$

A simple calculation shows that

$$\gamma_{AB} = \gamma_{BC} = \gamma_{CD} = \gamma_{CE} = 8,$$

hence $\gamma$ has the same value for all edges of the graph. Such a graph is called *balanced.*

More generally, we may consider a fairly arbitrary graph and investigate quantities like the above $\delta$'s and $\gamma$'s. In this paper we present some of our findings. Section 1.2 contains graph-theoretic preliminaries; in Section 1.3 we give rigorous definitions of the random walks and of most of the quantities we are interested in. In Section 3 we treat the special subject of balanced graphs.

The origin of our work was the question: "How long will it take to reach the goal of a maze if I move at random through the maze?" In the obvious model suggested by the above example, the sequence of successive vertices is of course a Markov chain with a finite state space. Although detailed results for such chains are available in the literature, we believe that the specific findings collected in this paper may be of interest.

## *1.2. Graphs*

Let $G$ be a finite, *connected*, undirected graph without loops and, unless stated otherwise, without multiple edges. Whenever we use the word graph, we tacitly assume that it has all the above properties; we use the word multigraph if all the conditions are fulfilled with the last as a possible exception. Often we write $(X, \Gamma)$ instead of $G$, denoting by $X$ the set of vertices of $G$ and by $\Gamma$ the set of edges. By $\Gamma$ a multivalued mapping from $X$ into $X$ (to be denoted by $\Gamma$ as well) is defined in such a way that, if $x \in X$, $\Gamma x$ is the set of all $y \in X$ adjacent to $x$.

As usual, $\bigcup_{x \in B} \Gamma x$ will be denoted by $\Gamma B$ if $B$ is a subset of $X$. Since $G$ has no multiple edges, $G$ is uniquely determined by $X$ and the mapping $\Gamma$. Usually, we shall denote $|X|$ by $n$ and $|\Gamma|$ by $r$. The valency (or degree) $v_x$ of $x \in X$ is by definition equal to $|\Gamma x|$; the index $x$ will be dropped frequently when no confusion is possible, e.g. when all valencies are equal. Note that $\Sigma_{x \in X} v_x = 2r$.

If $Y$ is a subset of $X$, then the subgraph generated by $Y$ is by definition the subgraph with vertex set $Y$, hence the graph with vertex set $Y$ and which has as edges all edges of $X$ with both end points in $Y$.

The vertices of $G$ are usually supposed to be labeled $1, 2, ..., n$ or $0, 1, ..., n-1$.

The *adjacency matrix* of $G$ is the $n \times n$ matrix $A = [A_{ij}]$, with $A_{ij} = 1$ if $i \in \Gamma_j$ and $A_{ij} = 0$ otherwise. Note that $A$ is a symmetric matrix with zeros on the diagonal and that

$$\sum_i A_{ij} = \sum_i A_{ji} = v_j.$$

A *bridge* of $G = (X, \Gamma)$ is an edge $e$ such that $(X, \Gamma - \{e\})$ is not connected. A *cut-point* of $G$ is a vertex $x$ such that the subgraph generated by $X - \{x\}$ is not connected. A *block* is a graph without cutpoints.

Let $G = (X, \Gamma)$ and $H = (Y, \Delta)$ be graphs, and $x \in X$, $y \in Y$. The *treewise join* of $G$ and $H$ (at $x$ and $y$) is the graph obtained by just mutually identifying $x$ and $y$.

The *straight distance* $m_{xy}$ of two vertices $x$ and $y$ of a graph $G$ is the smallest number $k$ such that $x \in \Gamma^k y$ (where $\Gamma^k$ is the $k^{\text{th}}$ iterate of $\Gamma$).

Some special graphs, which we need in the sequel, are the following:

(a) The complete graph on $n$ vertices, $K_n$, which is the graph with $\binom{n}{2}$ edges.

(b) The complete bipartite graph $K_{m,n}$, where $X = A + B$ with $A$ and $B$ disjoint, $|A| = m$, $|B| = n$, and

$$\Gamma = \{\{a, b\} \mid a \in A, b \in B\}.$$

(c) The cyclic graph $C_n$, where $X = \{0, 1, 2, ..., n-1\}$, and

$$\Gamma = \{\{i, i+1\} \mid i = 0, ..., n-2\} \cup \{\{0, n-1\}\}$$

(d) The linear graph $L_n$, where $X = \{0, 1, 2, ..., n-1\}$, and

$$\Gamma = \{\{i, i+1\} \mid i = 0, ..., n-2\}.$$

If $G$ is a graph, then $G^k$, the *k-fold multigraph corresponding to G* is defined as the multigraph which can be obtained from $G$ by replacing each edge with $k$ edges ($k \geqslant 1$, $G^1 = G$). See Fig. 1.2.

If $G$ and $H$ are graphs and $x$ and $y$ are two distinct equivalent vertices of $H$, then we denote by $\langle x, y, H\rangle \to G^k$ (or if $x$, $y$ are fixed, by $H \to G^k$) the graph which can be obtained from $G^k$ by replacing each edge of $G^k$ with a copy of $H$ in such a way that $x$ and $y$ are identified with $i$ and $j$ respectively, if $i$ and $j$ are neighbours in $G$. For an example, see Fig. 1.3. Note that $\langle 0, 3, L_4\rangle \to C_3^2$ can also be obtained as $\langle 0, 3, C_6\rangle \to C_3^1$.

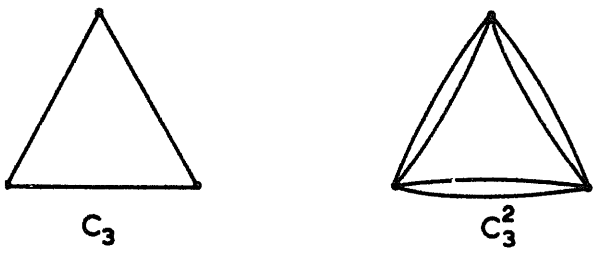


Fig. 1.2. A graph and its 2-fold multigraph.

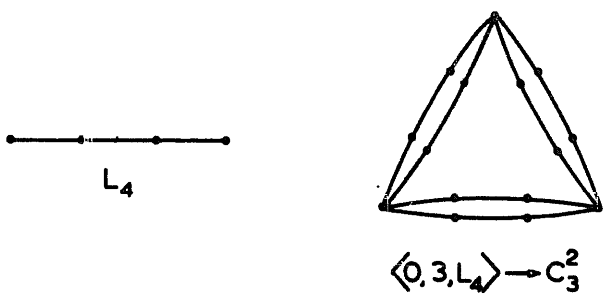


Fig. 1.3. An example of substitution.

## 1.3. Random walks and related quantities

Let $G = (X, \Gamma)$ be a graph; to avoid trivialities we assume $|X| > 1$. Let $i \in X$. A *random walk on G starting at i* is a random sequence $x_0, x_1, x_2, \ldots$ of vertices of $G$ such that:

(a) $\mathbf{P}[x_0 = i] = 1$,

(b) $\mathbf{P}[x_{m+1} = k \mid x_0 = i, x_1 = i_1, \ldots, x_{m-1} = i_{m-1}, x_m = j] =$
$= \mathbf{P}[x_{m+1} = k \mid x_m = j] = v_j^{-1}$ if $k \in \Gamma_j$, and $= 0$ otherwise.

Hence $x_0, x_1, x_2, \ldots$ is a Markov chain with stationary transition probabilities and with finite state space $X$. The chain is irreducible (since $G$ is connected) and either aperiodic or, if $G$ is bipartite, periodic with period 2.

Let $i$ and $j$ be elements of $X$. We define the stochastic variable $d_{ij}$ to be identically 0 if $i = j$. If $i \neq j$, then we say that $d_{ij}$ assumes the value $k$ if and only if in a random walk starting at $i$ we have

$$x_1 \neq j, \quad x_2 \neq j, \quad \ldots, \quad x_{k-1} \neq j, \quad x_k = j.$$

Hence $d_{ij}$ is the number of steps to reach $j$ from $i$ for the first time. The expectation of $d_{ij}$ is denoted by $\delta_{ij}$. From our assumptions it follows that every two states communicate, and, since $X$ is finite, it follows that $\delta_{ij} < \infty$.

If $i \in X$, then $e_i$ is the recurrence time of $i$, i.e. $e_i = k > 0$ if and only if

$$x_k = i, \quad x_{k-1} \neq i, \quad \ldots, \quad x_1 \neq i, \quad x_0 = i.$$

The expectation $E\{e_i\}$ is denoted by $\epsilon_i$. Note that $\epsilon_i < \infty$.

The quantities $\delta_{ij}$ and $\epsilon_i$ satisfy the two closely related recursions

$$\delta_{ij} = 1 + v_i^{-1} \sum_{k \in \Gamma i} \delta_{kj}, \quad i \neq j, \tag{1.1}$$

$$\epsilon_i = 1 + v_i^{-1} \sum_{k \in \Gamma i} \delta_{ki}. \tag{1.2}$$

These can be easily proved by considering the conditional expectations, given the outcome of the first step, in a random walk starting at $i$. In terms of matrices, (1.1) and (1.2) can be written as

$$\delta + \epsilon = E + P\delta,$$

where $\delta$ is the $n \times n$ matrix with entries $\delta_{ij}$, $\epsilon$ is the diagonal $n \times n$ matrix with $\epsilon_i$ on the diagonal, $E$ consists of 1's, and $P$ is the transition probability matrix. Note that $P$ can be written as $P = VA$, where $V$ is the diagonal matrix with $v_i^{-1}$ on the diagonal and $A$ is the adjacency matrix.

Several important results on $d_{ij}$ and $e_i$ can be obtained from equations like (1.1) by simple manipulations from linear algebra. For example, eq. (1.1) can be solved in terms of the so-called fundamental matrix [3, Theorem 4.4.7]. Also, according to [3, Theorem 4.4.4] we have $\epsilon_i \pi_i = 1$, a result to be explained and exploited presently. For further examples, see [3, chs. IV, V].

**Theorem 1.1.** *If $G = (X, \Gamma)$ and $i \in X$, then $\epsilon_i = 2r/v_i$.*

**Proof.** A finite irreducible Markov chain has a stationary distribution, say $\pi_i$ ($i = 1, 2, \ldots$), which is the unique solution of

$$\sum \pi_i p_{ij} = \pi_j, \quad j = 1, 2, \ldots, \qquad \sum \pi_i = 1,$$

where $p_{ij}$ are the transition probabilities. By substitution one easily verifies that

$$\pi_i = \tfrac{1}{2} v_i / r, \quad i = 1, 2, \ldots,$$

constitute a solution of the above system, and the result follows from $\epsilon_i = \pi_i^{-1}$. □

**Corollary 1.2.** *If* $v_j = v$ *for all* $j$, *then* $\epsilon_j = n$ *for all* $j$.

**Remark 1.3.** Theorem 1.1 and Corollary 1.2 are also true when $G$ is a multigraph; the proof is virtually the same.

**Proposition 1.4.** *If* $\{1\} \cup \Gamma 1 = \{2\} \cup \Gamma 2$, *then*

$$\delta_{12} = \delta_{21} = 2r/(v_1 + 1).$$

**Proof.** Apply (1.1) with $i = 1, j = 2$, and (1.2) with $i = 2$. □

If $i$ and $j$ are elements of $X$, the *commuting time* $\boldsymbol{c}_{ij}$ is defined as the number of steps in a random walk from $i$ to $j$ and back. Hence $\boldsymbol{c}_{ij}$ has the same distribution as $\boldsymbol{d}_{ij} + \boldsymbol{d}_{ji}$. The expectation $\mathrm{E}\{\boldsymbol{c}_{ij}\}$ is denoted by $\gamma_{ij}$. If $i$ and $j$ are joined by an edge $e$ of $\Gamma$, then $\gamma_{ij}$ may be alternatively denoted by $\gamma_e$.

Consider a random walk which starts at $i$ and which stops as soon as $k$ is reached. The number of times the vertex $j$ is left during this random walk will be denoted by $\boldsymbol{b}_{ijk}$, and its expectation by $\beta_{ijk}$. We define $\boldsymbol{b}_{iji} \equiv 0$ and $\boldsymbol{b}_{ijj} \equiv 0$ for all $i, j$. The quantities $\boldsymbol{b}_{ijk}$ satisfy the recurrence relations

$$\beta_{ijk} = \delta_i^j + v_i^{-1} \sum_{s \in \Gamma i} \beta_{sjk}, \quad i \neq k, \tag{1.3}$$

where $\delta_i^j$ is the Kronecker delta. The proof of (1.3) is similar to that of (1.1) and (1.2).

In terms of matrices, (1.3) reads

$$\beta_k = I + P_k \beta_k, \tag{1.4}$$

where $\beta_k$ is the $(n-1) \times (n-1)$ matrix with entries $\beta_{ijk}$ $(i \neq k, j \neq k)$, $I$ is the identity matrix and $P_k$ the matrix obtained from $P$ by deleting the $k^{\text{th}}$ row and $k^{\text{th}}$ column. The formal solution of (1.4) is

$$\beta_k = (I - P_k)^{-1} = \sum_{t=0}^{\infty} P_k^t. \tag{1.5}$$

For a proof of the existence of the inverse, the convergence of the series and several other results we refer to [3, ch. III]. From the definitions it is clear that $\boldsymbol{d}_{ij}$ has the same distribution as

$$\sum_{k \in X} \boldsymbol{b}_{ikj},$$

and hence

$$\delta_{ij} = \sum_{k \in X} \beta_{ikj}. \tag{1.7}$$

This breakdown of $\delta_{ij}$ coincides with [3, Theorem 3.3.5].

**Example 1.5.** Let $G = C_n$, labeled in the usual way. Let $0 < j < n$, and let the vertex 0 alternatively be denoted by $n$. Then (1.3) takes the form

$$\beta_{ijn} = \tfrac{1}{2}\, \beta_{i+1,j,n} + \tfrac{1}{2}\beta_{i-1,j,n} \quad \text{for } 0 \neq i \neq j, \tag{1.8}$$

$$\beta_{jjn} = 1 + \tfrac{1}{2}\beta_{j-1,j,n} + \tfrac{1}{2}\beta_{j+1,j,n}. \tag{1.9}$$

From these relations and from the boundary conditions $\beta_{0jn} = \beta_{njn} = 0$, we obtain

$$\beta_{ijn} = 2i(n-j)/n, \quad 0 \leqslant i \leqslant j,$$

$$\beta_{ijn} = 2j(n-i)/n, \quad j \leqslant i \leqslant n.$$

Or, in one formula:

$$\beta_{ijn} = 2\{\min(i, j) - ij/n\}. \tag{1.10}$$

A remarkable property is the symmetry of the $\beta_{ijn}$, viz. $\beta_{ijn} = \beta_{jin}$. This is not a specific property of $C_n$, as is shown by the following proposition.

**Proposition 1.6.** *$\beta_{ikj}/v_k = \beta_{kij}/v_i$ for any graph $G$.*

**Proof.** By (1.5), $\beta_j^{\mathrm{T}} = (I - P_j^{\mathrm{T}})^{-1}$. Furthermore, if $V_j$ is obtained from $V$, as $P_j$ from $P$, by deleting the $j^{\text{th}}$ row and column, then $P_j^{\mathrm{T}} = V_j^{-1} P_j V_j$. Hence $\beta_j^{\mathrm{T}} = V_j^{-1} \beta_j V_j$, or $V_j \beta_j^{\mathrm{T}} = \beta_j V_j$ and so $\beta_{ikj}/v_k = \beta_{kij}/v_i$ (at first only for $i \neq j$ and $k \neq j$). □

### *1.4. Time reversal*

In general, if $P$ is the transition matrix of a finite ergodic Markov chain, the reverse Markov chain is a Markov chain with transition matrix $\hat{P}$ given by $\hat{P} = \epsilon P^{\mathrm{T}} \epsilon^{-1}$. A Markov chain is called *reversible* if $P = \hat{P}$ [3, Definitions 5.3.1, 5.3.2]. Clearly $P = \hat{P}$ here, since $P = V^{-1} P^{\mathrm{T}} V$ and $\epsilon = 2r\, V^{-1}$ by Theorem 1.1. Essentially from this fact, Proposition 1.6 was derived, and indeed this very proposition can be used to transform the forward recursion (1.3)

$$\beta_{ikj} = \delta_i^k + v_i^{-1} \sum_{s \in \Gamma i} \beta_{skj}, \quad i \neq j,$$

into a backward recursion:

$$\beta_{ikj} = \delta_i^k + \sum_{s \in \Gamma k} (\beta_{isj}/v_s), \quad k \neq j. \tag{1.11}$$

A more direct proof follows from (1.4), noting that, since $\beta_j$ is invertible, $\beta_j = I + P_j\beta_j$ iff $\beta_j = I + \beta_j P_j$.

Formula (1.11) has a nice interpretation, which may also serve as an alternative proof. Namely, $\beta_{ikj}/v_k$ is the expected number of times a directed edge $(k, x)$ is traversed in a random walk from $i$ to $j$. Hence

$$\sum_{s \in \Gamma k} (\beta_{isj}/v_s)$$

is the expected number of arrivals at $k$, while $\beta_{ikj}$ is, by definition, the expected number of departures from $k$.

By considering the final step of a random walk from $i$ to $j$, the same interpretation yields the following supplement to (1.11):

$$1 = \sum_{s \in \Gamma j} (\beta_{isj}/v_s), \quad i \neq j. \tag{1.12}$$

Or, (1.11) and (1.12) combined:

$$\delta_j^k + \beta_{ikj} = \delta_i^k + \sum_{s \in \Gamma k} (\beta_{isj}/v_s) \tag{1.13}$$

for all $i, j, k$.

## 2. Results

### 2.1. *The stochastic triangle inequality and other general theorems*

**Definition 2.1.** If $x$ and $y$ are two random variables with distribution functions $F$ and $H$, respectively, then we say that $x$ is *stochastically less than or equal to* $y$ (notation: $x \leqslant^s y$) iff $F(x) \geqslant H(x)$ for all $x$.

Clearly, if $x \leqslant^s y$ and $y \leqslant^s x$, then $x \triangleq y$.

Let $G = \{X, \Gamma\}$ be a graph.

**Definition 2.2.** If $i,j,k \in X$, we say that *k lies between i* and *j* if $k = i$ or $k = j$ or *every* path from $i$ to $j$ contains $k$. (Hence $k$ is a cut point in the last case).

**Theorem 2.3.**
(a) $d_{ij} \overset{s}{\leqslant} d_{ik} + d_{kj}$
(b) $d_{ij} \triangleq d_{ik} + d_{kj}$ *iff k lies between i and j.*
(c) $e_i \overset{s}{\leqslant} c_{ik}$.
(d) $e_i \triangleq c_{ik}$ *iff* $\Gamma i = \{k\}$.

A proof can be given with the aid of Chung's decomposition theorem [1, p. 46]. □

From the well-known result

$$\mathrm{E}\{x\} = \int_0^\infty \{1 - F(x)\}\, \mathrm{d}x$$

for an arbitrary non-negative random variable $x$ with distribution function $F(x)$, we have the following corollaries.

**Corollary 2.4.**
(a) $\delta_{ij} \leqslant \delta_{ik} + \delta_{kj}$.
(b) $\delta_{ij} = \delta_{ik} + \delta_{kj}$ *iff k lies between i and j.*
(c) $\epsilon_i \leqslant \gamma_{ik}$.
(d) $\epsilon_i = \gamma_{ik}$ *iff* $\Gamma i = \{k\}$.

From (a) we have $\gamma_{ij} \leqslant \gamma_{ik} + \gamma_{kj}$, and since $\gamma$ is symmetric, a graph is a metric space with respect to the distance $\gamma$.

**Theorem 2.5.**
(a) $\Sigma_{\{i,j\} \in \Gamma} (\beta_{ikj} + \beta_{jki})/v_k = n-1$.
(b) $\Sigma_{e \in \Gamma}\, \gamma_e = 2r(n-1)$.

**Proof.** According to Proposition 1.6 and formula (1.12),

$$1 - \delta_j^k = \sum_{i \in \Gamma j} (\beta_{ikj}/v_k).$$

From this, (a) follows by summation over $j$, and then (b) from (a) by summation over $k$. □

**Proposition 2.6.** *All quantities which are derived from the Markov chain related to G, such as $\gamma$, $\delta$, $\epsilon$, etc., are the same for G and $G^k$.*

**Proof.** Note that $P^G = P^{G^k}$, the transition probabilities being the same. □

**Theorem 2.7.** *Suppose G and H are graphs, x and y are distinct equivalent vertices of H, and K is $\langle x, y, H\rangle \to G^k$. Let i and j be vertices of G, embedded in the natural way into K. Then*

$$\delta_{ij}^K = \delta_{xy}^H \ \delta_{ij}^G.$$

**Proof** (outline). Let $X$ be the vertex set of $G$, embedded in the natural way into K. Consider the random walk in $K$ starting at $i$, which stops as soon as $j$ is reached. Now for each realization $\{x_0 = i, x_1, x_2, ...\}$ of this random walk, we define an ***interval*** **(*of length a*)** as a subsequence of the form $\{x_s, x_{s+1}, x_{s+2}, ..., x_{s+a}\}$ which satisfies the following conditions:

(a) $x_s \in X$, $x_{s+a} \in X$, $x_s \neq x_{s+a}$;

(b) $x_p \in X$ with $s \leqslant p < s + a$ implies $x_p = x_s$.

Note that each realisation is *uniquely* partitioned into intervals. Let $N$ be the number of intervals in $\{x_0 = i, x_1, x_2, ...\}$, and $t_m$ the length of the $m^{\text{th}}$ interval ($m = 1, 2, ..., N$). For the corresponding random variables $N$ and $t_m$ we have

$$d_{ij} = t_1 + t_2 + ... + t_N,$$

and hence

$$\delta_{ij} = \mathbf{E}\{t_1 + ... + t_N\} = \mathbf{E}\,\mathbf{E}\{t_1 + ... + t_N \mid N = N\}$$

$$= \mathbf{E}\{t_1\}\ \mathbf{E}\{N\} = \mathbf{E}\{d_{xy}^H\}\ \mathbf{E}\{d_{ij}^G\} = \delta_{xy}^H\ \delta_{ij}^G,$$

where we have made use of the fact that $t_1$ and $d_{xy}^H$ have the same distribution, as well as $N$ and $d_{ij}^G$. □

## 2.2. Decomposition theorems

**Lemma 2.8.** *Let $G = (X, \Gamma)$ be a multigraph with a cut point p. Let B be a subset of $X - \{p\}$ such that $B \cup \Gamma B = B \cup \{p\}$. (Hence B contains all vertices of some components of $X - \{p\}$). Let k be the number of edges from p to B, and $r_0$ the number of edges of G which have no end vertex in B. Then*

$$\delta_{pB} = 1 + 2r_0/k.$$

**Remark 2.9.** A greek letter with a graph symbol as a superscript indicates a conditional expectation. For example, when $H$ is a partial subgraph of $G$, and $p$ is a vertex of $H$, then $\epsilon_p^H$ is the conditional expectation of the recurrence time to $p$, under the condition that the random walk is restricted to vertices of $H$.

**Proof of Lemma 2.8.** Let $H$ be the sub-multigraph of $G$ generated by $X - B$. Then according to Theorem 1.1 we have

$$\epsilon_p^H = 2r_0/(v_p - k),$$

where $v_p$ is the (total) valency of $p$. Hence

$$\delta_{pB} = v_p^{-1}\{k + (v_p - k)(\epsilon_p^H + \delta_{pB})\}$$

$$= v_p^{-1}\{k + 2r_0\} + (1 - k/v_p)\delta_{pB}.$$

Solving for $\delta_{pB}$, we obtain the desired result. □

Lemma 2.8 is more powerful, as we shall see presently, when we combine it with imploding, to use a graph-theoretic term, or lumping of states in Markov chain terminology.

**Lemma 2.10.** *Let $G = (X, \Gamma)$ be a multigraph, and let $B$ be a non-trivial subset of $X$. Let*

$$C = \Gamma B - B = \{c_1, c_2, ..., c_m\}.$$

*Suppose $\delta(c_j, B) = \delta(C, B)$ does not depend on $j$. Further suppose that there are no edges between any pair $(c_i, c_j)$. Let there be exactly $k$ edges from each $c_j$ to $B$, and $r_0$ edges without end-points in $B$. Then*

$$\delta(C, B) = 1 + 2r_0/(km).$$

**Remark 2.11.** In our applications, the condition on $\delta(c_j, B)$ is usually trivially true for symmetry reasons.

**Proof of Lemma 2.10.** Consider the auxiliary multigraph $(X', \Gamma')$ obtained from $(X, \Gamma)$ by mutually identifying all $c_j$. The conditions of the lemma ensure that $\delta(C, B)$ does not change. By applying Lemma 2.8, we immediately obtain the desired result. □

**Remark 2.12.** The restriction to graphs without edges between $c_i$ and $c_j$ has been made to avoid loops in the multigraph $(X', \Gamma')$.

## 2.3. Applications

*2.3.1. Distances in the cube graph.* The $D$-dimensional cube graph may be defined as follows. The vertex set $X$ is the set of all $2^D$ binary sequences of length $D$. Two vertices are connected by an edge iff the corresponding sequences differ in exactly one place. Hence the valency of a vertex is $D$, and the total number of edges is $D\ 2^{D-1}$.

For technical reasons, suppose that $X$ is a subset of $\mathbf{R}^D$ equipped with the norm $\|\cdot\|$ given by

$$\|x\| = \sum_{i=1}^{D} |x_i|$$

for $x = (x_1, ..., x_D) \in \mathbf{R}^D$. (Then $x,y \in X$ are connected by an edge iff $\|x-y\| = 1$.) Set

$$S_m = \{x \in X \mid \|x\| = m\}.$$

Since every $x \in S_m$ is connected with $S_{m-1}$ by $m$ edges, and since, for symmetry reasons, $\delta(x, S_{m-1})$ does not depend on $x$ if $x$ ranges over $S_m$, Lemma 2.10 can be applied, yielding

$$\delta(S_m, S_{m-1}) = \binom{D-1}{m-1}^{-1} \sum_{j=m}^{D} \binom{D}{j}.$$

Now by Corollary 2.4(b) we have

$$\delta(S_k, S_0) = \sum_{m=1}^{k} \delta(S_m, S_{m-1}),$$

so that in general, by an obvious symmetry argument,

$$\delta_{xy} = \sum_{m=1}^{k} \binom{D-1}{m-1}^{-1} \sum_{j=m}^{D} \binom{D}{j} \quad \text{if } \|x-y\| = k.$$

In particular we find that the maximum value of $\delta_{xy}$, where $x$ and $y$ run through the vertex set $X$, is given by

$$\delta_{\max} = \sum_{m=1}^{D} \binom{D-1}{m-1}^{-1} \sum_{j=m}^{D} \binom{D}{j},$$

which can be reduced to

$$\delta_{\max} = 2^{D-1} \sum_{m=0}^{D-1} \binom{D-1}{m}^{-1}.$$

The derivation constitutes a nice exercise with binomial coefficients.

We note that for $D$ large,

$$\delta_{\max} \sim 2^D (1 + D^{-1}),$$

whereas

$$\delta_{x,y} = 2^D - 1 \quad \text{if } \|x-y\| = 1.$$

*2.3.2. Collecting times on the cyclic graph.* Let $G = (X, \Gamma)$ be a graph and let $i \in X$. Consider the random walk which starts at $i$ and which stops as soon as all vertices have been visited at least once. The number of steps in such a random walk is called the *collecting time starting at i,* and is denoted by $x_i$. $\mathrm{E}\{x_i\}$ is denoted by $\xi_i$ and $\max_{i \in X} \xi_i$ by $\xi$.

We show that $\xi = \binom{n}{2}$ when $G = C_n$. When we have just added the $k^{\text{th}}$ (new) vertex to our collection, the expected number $\eta_k$ of steps required to obtain the $k + 1^{\text{st}}$ vertex can be found by applying Lemma 2.10 with $B$ equal to the set of vertices not in the collection (note that the starting point is indeed adjacent to $B$):

$$\eta_k = 1 + 2(k-1)/(1 \cdot 2) = k,$$

and hence

$$\xi = \eta_1 + \eta_2 + \ldots + \eta_{n-1} = \binom{n}{2}.$$

## 2.4. Tree-wise joins

Let $G = (X, \Gamma)$, with $X = \{x_1, x_2, \ldots, x_n\}$, and let $H_j = (Y_j, \Delta_j)$, with $y_j \in Y_j$ $(j = 1, 2, \ldots, n)$. The *tree-wise join* (or just the *join*) of $G$ and $H_1, \ldots, H_n$ is defined as the graph $H$ which arises by identifying $x_j$ and $y_j$ $(j = 1, 2, \ldots, n)$. Note that we do not exclude the possibility that one or several $H_i$ consist of a single point. An example is shown in Fig. 2.1.

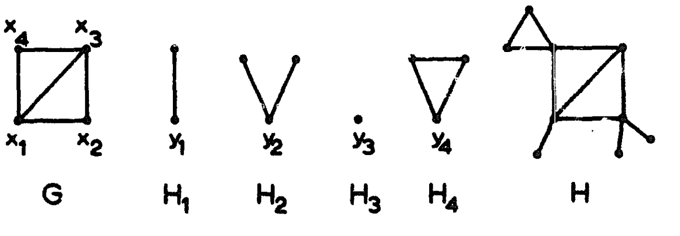


Fig. 2.1. The join $H$ of $G$ and $H_1$, $H_2$, $H_3$, $H_4$.

**Theorem 2.13.** *Let $G = (X, \Gamma)$ be a multigraph. Let $H$ be the join of $G$ and $H_j = (Y_j, \Delta_j)$, with $|\Delta_j| = r_j$ $(j = 1, 2, ..., n)$. Let $i$ and $k$ be elements of $X$. Then*

$$\delta_{ik}^H = \delta_{ik}^G + 2 \sum_{j\in X} (r_j \beta_{ijk}^G / v_j^G), \tag{2.1}$$

*or, equivalently,*

$$\delta_{ik}^H = \sum_{j\in X} \beta_{ijk}^G (1 + 2r_j/v_j^G). \tag{2.2}$$

**Proof** (outline). Consider a random walk on the graph $H$. Part of it will lie in $G$; its expected length is $\delta_{ik}^G$. Besides, there will be a number of departures from $j$ to a vertex of $G - \{j\}$; the expectation of this number is $\beta_{ijk}^G$. Each such departure from $j$ is preceded by a random walk in $H_j$, possibly of length 0; the expected length of one such random walk in $H_j$ follows at once from Lemma 2.8: it is equal to

$$\delta_{j, G-\{j\}} - 1 = 2r_j/v_j^G.$$

We subtract 1 because the step from $j$ to $G - \{j\}$ is part of the walk within $G$ and has been accounted for in the number $\delta_{ik}^G$. □

*2.5. The attraction law for a tree*

Let $i$ and $j$ be vertices of a tree $H$ with straight distance $m_{ij} = n$. Denote the successive vertices between $i$ and $j$ inclusive by $0, 1, ..., n$. Then $0, 1, ..., n$ generate a subgraph $G$ of $H$ isomorphic to $L_{n+1}$. Moreover, $H$ is the tree-wise join of $G$ and trees $H_k$ $(0 \leqslant k \leqslant n)$, attached to $G$ at the vertex $k$, and with, say, $n_k$ vertices (including $k$). Hence by Theorem 2.13 we have

$$\delta_{0n}^H = \sum_{k=0}^{n} \beta_{0kn}^G \{1 + 2(n_k - 1)/v_k^G\}.$$

Now $v_k^G = 2$ $(k = 1, ..., n-1)$, and $v_0^G = v_n^G = 1$, while a straightforward calculation shows that

$$\beta_{0kn}^{L_{n+1}} = 2(n-k), \quad k > 0, \qquad \beta_{00n}^{L_{n+1}} = n.$$

Hence

$$\tfrac{1}{2}(\delta_{0n}^H + m_{0n}^H) = \sum_{k=0}^{n} n_k (n-k). \tag{2.3}$$

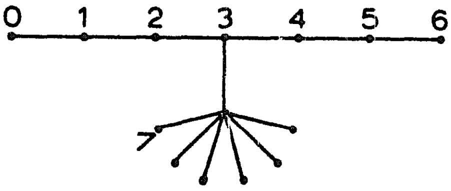


Fig. 2.2. A tree in which $\delta_{max}$ and the diameter are attained for different pairs of vertices.

This formula has a nice interpretation. By giving all vertices of $H$ mass 1 and identifying the vertices of $G$ with the corresponding points $0, 1, ..., n$ on the real line followed by concentrating the whole mass of $H_k$ in $k$ $(0 \leq k \leq n)$, a mass distribution on the line is obtained. The (euclidean) distance of its centre of gravity to the point $n$ is proportional to the right-hand side of (2.3), thereby providing an interpretation of this formula as a sort of attraction law.

With the aid of (2.3), distances $\delta_{ij}$ in a tree can be quickly calculated. In the tree of Fig. 2.2 we find $\delta_{0,6} = 78$ and $\delta_{7,6} = 83$. This shows that in a tree, $\delta_{max}$ is not necessarily attained for two vertices of which the straight distance is the diameter. (For the behaviour of $\gamma$ in this respect see Section 3.2).

*2.6. The quantities $\theta_{ij}$*

We have already met the quantity $\beta_{ijk}/v_j$ several times. From its interpretation given in Section 1.4 it follows that the symmetrized quantity

$$\theta_{ijk} = (\beta_{ijk} + \beta_{kji})/v_j$$

may be interpreted as the expected number the step $j \to s$ is made in a random walk from $i$ to $k$ and back (where $s$ is a fixed element of $\Gamma j$). This quantity, $\theta_{ijk}$, will be of crucial importance in the sequel. Theorem 2.5(a) states that

$$\sum_{\{i,j\} \in \Gamma} \theta_{ikj} = n-1,$$

independent of $k$. We now prove the following much stronger result.

**Theorem 2.14.** *In each $G = (X, \Gamma)$ with $i,j,k \in X$, $\theta_{ikj}$ does not depend on $k$.*

**Proof.** If one symmetrizes formula (1.13) with respect to $i$ and $j$, one obtains

$$v_k \theta_{ikj} = \sum_{s \in \Gamma k} \theta_{isj} = \sum_{s \in X} (v_s \theta_{isj}) p_{sk}, \quad k = 1, 2, \ldots .$$

In other words, for fixed $i$ and $j$ the quantities $v_k \theta_{ikj}$ satisfy the stationary state equations. Hence

$$v_k \theta_{ikj} = \lambda_{ij} \pi_k ,$$

where $\lambda_{ij}$ is a constant depending on $i$ and $j$ only. The desired result now follows from Theorem 1.1. □

From now on we write $\theta_{ij}$ for $\theta_{ikj}$.

**Corollary 2.15.** $\theta_{ij} = \beta_{iij}/v_i = \beta_{jji}/v_j$.

**Corollary 2.16.** *In each graph $G = (X, \Gamma)$ with $i, k \in X$,*

$$\theta_{ik} = \tfrac{1}{2} \gamma_{ik}/r.$$

**Proof.** From (2.2) it follows by interchanging $i$ and $k$ and adding that

$$\gamma_{ik}^H = \sum_j (2r_j + v_j^G)\, \theta_{ik}^G = 2r^H\, \theta_{ik}^G .$$

Since we may choose $H = G$, the corollary follows. □

**Corollary 2.17.**
(a) $\theta_{ik} + \theta_{kj} = \theta_{ij} + 2\beta_{ijk}/v_j$.
(b) $\gamma_{ik} + \gamma_{kj} = \gamma_{ij} + 2(2r/v_j)\, \beta_{ijk}$.

**Proof.** Apply Proposition 1.6. □

**Remark 2.18.** In view of Corollary 2.16, (a) and (b) are equivalent. Both (a) and (b) can be viewed as triangle inequalities, in which the deviation from equality is precisely given (namely by the last term). Compare [1, §11, Theorem 3].

Using (1.7) (and Theorem 1.1) one easily obtains the following corollary from Corollary 2.17 (after multiplication by $v_j$ and summation over $j$).

**Corollary 2.19.** $\delta_{ki} - \delta_{ik} = \Sigma_j\, v_j(\theta_{ij} - \theta_{kj}) = \Sigma_j\, \pi_j(\gamma_{ij} - \gamma_{kj})$.

This shows among other things that the $\delta$-structure is determined by the $\gamma$-structure.

**Theorem 2.20.** *If $H$ is obtained by joining graphs $H_j$ to $G = (X, \Gamma)$, then for each pair $i,k \in X$,*

$$\theta_{ik}^G = \theta_{ik}^H .$$

**Proof.** From the proof of Corollary 2.16 we have $\theta_{ik}^G = \gamma_{ik}^H/(2r^H)$. By the same corollary applied to $H$ we have $\theta_{ik}^H = \gamma_{ik}^H/(2r^H)$, and the theorem follows. □

## 3. Balanced graphs

### *3.1. General theorems*

**Theorem 3.1.** *If $G = (X, \Gamma)$ and $j \in X$ has valency $v_j$ with $i,k,s, \ldots \in \Gamma j$, then:*

(a) $v_j(\theta_{ij} - \theta_{sj}) = \Sigma_{k \in \Gamma j}\, (\theta_{ik} - \theta_{sk})$;
(b) $v_j(-1 + \Sigma_{i \in \Gamma j}\, \theta_{ij}) = \Sigma_{\{i,k\} \subset \Gamma_j}\, \theta_{ik}$.

**Proof.** Corollary 2.17(a) states that

$$\theta_{ij} + \theta_{jk} = \theta_{ik} + 2\beta_{ikj}/v_k .$$

Hence, by summation over all $k \in \Gamma j$, using formula (1.12):

$$v_j\theta_{ij} + \sum_{k \in \Gamma j} \theta_{kj} = \sum_{k \in \Gamma j} \theta_{ik} + 2.$$

This holds mutatis mutandis for any $s \in \Gamma j$ substituted for $i$, and (a) follows by subtraction, while (b) follows by summation over $i \in \Gamma j$. □

The result in (b) is less deep than the one in (a). In fact, (b) can be obtained using only the basic recursions (1.1) and (1.2) and the relation $\gamma_e = 2r\theta_e$. We leave the details as an exercise.

**Corollary 3.2.** *If $G = (X, \Gamma)$ and $j \in X$ has valency 2 with $\Gamma j = \{i, k\}$, then*

(a) $\theta_{ij} = \theta_{kj}$;
(b) $\theta_{ik} = 4\theta_{ij} - 2$.

**Proof.** Take $v_j = 2$ in Theorem 3.1. □

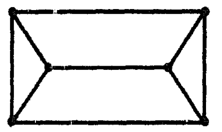


Fig. 3.1. A vertex-equivalent unbalanced graph.

**Definition 3.3.** A multigraph $G = (X, \Gamma)$ is called *balanced* when $\gamma_e$ (or equivalently $\theta_e$) has the same value for all $e \in \Gamma$.

Rather trivial examples of balanced graphs are edge-equivalent graphs like $K_n$, $C_n$ and $K_{m,n}$. Other examples will be given further on. An example of a *vertex*-equivalent graph which is not balanced is shown in Fig. 3.1. A direct calculation shows that $\gamma_a = \frac{54}{5}$, $\gamma_b = \frac{48}{5}$.

**Remark 3.4.** Although we consider connected graphs only, it seems natural to define $\delta_{ij} = \infty$ if $i$ and $j$ belong to different components of a disconnected graph. For disconnected graphs it does make a difference whether we require $\gamma_e$ or $\theta_e$ to be constant in order that the graph be balanced; $\theta_e$ (which gives the wider class) seems to be the better choice.

*3.2. Trees*

From Theorem 2.5 it follows that if a graph is balanced, then for all $e \in \Gamma$,

$$\theta_e = (n-1)/r, \qquad \gamma_e = 2(n-1)$$

(we do not need Corollary 2.16 for that). Note that Theorem 2.5 states that for any graph the *mean* of the $\theta_e$ is equal to $(n-1)/r$, and the mean of the $\gamma_e$ to $2(n-1)$.

**Proposition 3.5.** *Every tree is balanced.*

**Proof.** Let $e$ be an edge, with end points $p$ and $q$, of a tree. Since $e$ is a bridge, we may apply Lemma 2.8 with $k = 1$, and we obtain $\gamma_e = 2r = 2(n-1)$. □

**Proposition 3.6.** *A multigraph with a bridge is balanced iff the multigraph is a tree.*

**Proof.** Let $e$ be a bridge in a balanced multigraph $G$. Then $\gamma_e = 2r$, from the above proof, while on the other hand $\gamma_e = 2(n-1)$ since $G$ is balanced. Hence $n-1 = r$, and the "only if" part of the proposition follows from the connectedness of $G$. The "if" part is Proposition 3.5. □

*3.3. Tree-wise joins of balanced graphs*

If we join two graphs $G = (X, \Gamma)$ and $G' = (X', \Gamma')$ to form a graph $H$, then, according to Theorem 2.20,

$$\theta_e^H = \theta_e^G, \qquad \theta_{e'}^H = \theta_{e'}^{G'},$$

if $e \in \Gamma$ and $e' \in \Gamma'$. Hence a necessary and sufficient condition in order that the join of two graphs $G$ and $G'$ be balanced is that both $G$ and $G'$ are balanced with

$$\theta_e^G = \theta_{e'}^{G'}.$$

Since this is obviously also true for an arbitrary number of graphs, we have the following theorem.

**Theorem 3.7.** *The tree-wise join of graphs $G_1, G_2, ..., G_N$ is balanced iff all $G_i$ are balanced and have the same value of $\theta_e$.*

In particular, we may join *identical* balanced graphs, and the result will be balanced. Two *different* edge-equivalent graphs with the same value of $\theta_e$ are e.g. $K_m$ and $K_{m-1,m}$. Note that Proposition 3.5 is a special case of Theorem 3.7.

*3.4. A valency criterion*

Let $G = (X, \Gamma)$, $|X| = n$, $|\Gamma| = r$ and $v_0 = \min_{x \in X} v_x$.

**Proposition 3.8.** *If $2r \geqslant v_0\ (2n - v_0 - 1)$, then $G$ is not balanced unless $G$ is a tree or $G = K_n$.*

In particular (take $v_0 = n-2$), the graph obtainable from $K_n$ $(n > 3)$ by deleting one edge is not balanced.

To prove the proposition we need a lemma which is not without interest on its own.

**Lemma 3.9.** *If* $s \in \Gamma x$, *then* $\delta_{xs} \geqslant v_0 + (v_x - v_0)/v_x$.

**Proof.** The case $v_x = 1$ is trivial, so suppose $v_x \geqslant 2$. Fix $s \in \Gamma x$. Set

$$\alpha = \min\{\delta_{ks} \mid k \in \Gamma s, v_k \neq 1\}.$$

Let $k_1 \in \Gamma s$ satisfy $\delta_{k_1,s} = \alpha$. Then

$$\delta_{xs} = 1 + v_x^{-1} \sum_{\substack{k \in \Gamma x \\ k \neq s}} \delta_{ks} \geqslant 1 + v_x^{-1}(v_x - 1)\alpha.$$

Now $\alpha > 1$, so $v_{k_1} \neq 1$. Thus also

$$\alpha = \delta_{k_1 s} \geqslant 1 + v_{k_1}^{-1}(v_{k_1} - 1)\alpha \geqslant 1 + v_0^{-1}(v_0 - 1)\alpha.$$

Hence $\alpha \geqslant v_0$, and consequently

$$\delta_{xs} \geqslant v_0 + v_x^{-1}(v_x - v_0). \ \square$$

**Proof of Proposition 3.8.** Choose $x$ such that $v_x = v_0$. Let $\hat{\gamma}_e$ be the mean value of $\gamma_{xs}$ with $s$ ranging over $\Gamma x$. Then clearly

$$\hat{\gamma}_e = v_0^{-1} \sum_{s \in \Gamma x} (\delta_{sx} + \delta_{xs})$$

$$= 1 + v_0^{-1} \sum_{s \in \Gamma x} \delta_{sx} + v_0^{-1} \sum_{s \in \Gamma x} (\delta_{xs} - 1).$$

Thus, according to Lemma 3.9,

$$\hat{\gamma}_e \geqslant \epsilon_x + v_0 - 1 = v_0^{-1}\{2r + v_0(v_0 - 1)\}.$$

Hence, if $G$ is balanced, then

$$2r \leqslant v_0(2n - v_0 - 1).$$

This can be sharpened if we suppose in addition that $v_0 \neq 1$ and not all valencies are the same. We then choose $x$ such that $v_x = v_0$ and such that a vertex $y \in \Gamma x$ exists with $v_y > v_0$. Re-examining the proof of the lemma, one readily finds out that in this case

$$\delta_{xs} > v_0 + v_x^{-1}(v_x - v_0),$$

and that consequently

$$2r < v_0(2n - v_0 - 1),$$

as required. Note, however, that if $G$ is balanced and $2r = v_0(2n - v_0 - 1)$, and $G$ is neither a tree nor $K_n$, then it follows that $v_0 \neq 1$ and that not all valencies are the same. □

### *3.5 Substitution of linear graphs*

If we substitute linear graphs into a graph $G$, or more generally, place some extra vertices on some of its edges, we obtain a special sort of "linear" subgraphs of $G$, which we call segments.

**Definition 3.10.** Let $G = (X, \Gamma)$ be a graph. If

$$Y = \{s_0, s_1, ..., s_p\} \subset X$$

with all $s_j$ distinct, and $\Gamma s_k = \{s_{k-1}, s_{k+1}\}$ for all $k$ with $0 < k < p$, and

$$\Delta = \{f_k \mid 1 \leqslant k \leqslant p\},$$

where $f_k$ is the edge with end points $s_{k-1}$ and $s_k$, then $Y$ and $\Delta$ constitute a partial subgraph $S = (Y, \Delta)$ of $G$ called a *segment of $G$ of length $p$.*

Note that $S$ is the subgraph generated by $Y$ if $p = 1$ or $s_0 \notin \Gamma s_p$. Furthermore, any edge of $G$, together with its two end points, constitute a segment (obviously of length 1).

If $S$ is a segment, then, in view of Corollary 3.2, $\theta_e$ does not depend on the choice of $e$ from among the edges of $S$. Hence, if $G$ is edge equivalent, then $\langle 0, n, L_{n+1}\rangle \to G^k$ is balanced.

An example of a graph which is balanced by this remark is $\langle 0, n, L_{n+1}\rangle \to L_2^k$ (see Fig. 3.2). The condition that $G$ be edge equivalent can be weakened, as is shown by the following theorem.

**Theorem 3.11.** *Let $G$ be a graph. Then $\langle 0, n, L_{n+1}\rangle \to G^k$ is balanced iff $G$ is balanced.*

The proof rests on a lemma which is of some interest on its own.

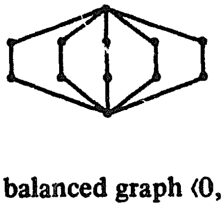

Fig. 3.2. The balanced graph $\langle 0, 3, L_4\rangle \to L_2^5$.

**Lemma 3.12.** *If S is a segment of a graph G and i, j, p, q are vertices of S, then*

$$m_{pq}^2(\theta_{ij} - m_{ij}) = m_{ij}^2(\theta_{pq} - m_{pq}),$$

*where $m_{ij}$ is the straight distance between i and j.*

*In particular, if $m_{ij} = k$ and e is an edge of S, then*

$$\theta_{ij} = k^2(\theta_e - 1) + k.$$

**Proof.** Denote $k + 1$ successive vertices of $S$ by 0, 1, ..., $k$ (i.e. $m_{i,i+1} = 1$ for all $0 \leq i < k$). We shall prove by induction on $k$ that $\theta_{0k} - k = k^2(\theta_{01} - 1)$. In view of Corollary 3.2 this is sufficient to prove the lemma. By Corollary 2.17(a), in combination with formula (1.12), we have

$$\theta_{0k} + \theta_{k,k+1} = \theta_{0,k+1} + 2 - 2v_{k-1}^{-1}\,\beta_{0,k-1,k}.$$

Since $\theta_{k,k+1} = \theta_{01}$ by Corollary 3.2(a), we will obtain an expression for $\theta_{0,k+1}$ in terms of $\theta_{0k}$ and $\theta_{01}$ if we determine $v_{k-1}^{-1}\beta_{0,k-1,k}$. We proceed as follows. Set

$$a_s = \beta_{0sk}/v_s, \quad 0 \leq s \leq k.$$

Then $a_0 = \theta_{0k}$ (Corollary 2.15), $a_k = 0$, and in-between $(0 < s < k)$ $a_s$ satisfies the familiar recursion

$$a_s = \tfrac{1}{2}a_{s+1} + \tfrac{1}{2}a_{s-1}$$

for the arithmetic sequence, as follows from formula (1.13). Hence

$$a_s = k^{-1}(k - s)\theta_{0k},$$

and consequently

$$\theta_{0,k+1} + 2 = \theta_{01} + k^{-1}(2 + k)\theta_{0k},$$

from which the proof of the induction step is easily derived. □

**Proof of Theorem 3.11.** We write $H = \langle 0, n, L_{n+1}\rangle \to G^k$. Let $e$ be an edge of $G$ with end points $x$ and $y$. In the construction of $H$ from $G$, a copy of $L_{n+1}$ is substituted for $e$; the result is a segment $S$ of $H$. Let $f$ be an arbitrary edge of $S$. We want to relate $\theta_e^G$ and $\theta_f^H$. We know that

$$\delta_{0n}^{L_{n+1}} = \delta_{n0}^{L_{n+1}} = n^2.$$

Hence it follows from Theorem 2.7, after the usual interchanging and adding that

$$\gamma^H_{xy} = n^2 \gamma^G_{xy}.$$

Or, since $r^H = kn\, r^G$,

$$\theta^H_{xy} = nk^{-1}\theta^G_{xy}.$$

Now Lemma 3.12 can be applied to the segment $S$ of $H$, and we obtain

$$\theta^G_e = k\{n(\theta^H_f - 1) + 1\}.$$

It now only takes a small step to complete the proof. □

*3.6. Various criteria*

We now further exploit Lemma 3.12. We start with a definition.

**Definition 3.13.** A vertex of a segment is called an *end vertex* (or *end point*) if it is incident with only one edge of the segment.

**Proposition 3.14.** *If within a graph $G$ two segments exist, both with the same end vertices but with distinct lengths $p$ and $q$, then $G$ is balanced only if $(n-1)\, r^{-1} = 1 - (p+q)^{-1}$.*

**Proof.** Apply Lemma 3.12. □

**Corollary 3.15.** *If within a graph $G$ three segments exist, all three with the same end vertices but with distinct lengths, then $G$ is not balanced.*

**Proposition 3.16.** *If $j$ is a vertex of $G$ with $v_j = 3$ and $i, k \in \Gamma j$ are the end points of a segment of length $p$, then $G$ is balanced only if*

$$(r - n + 1)\, r^{-1} = (p-2)\,(p^2 - 3)^{-1}.$$

Note that if $G$ satisfies the conditions of Proposition 3.16 with $p = 2$, then $G$ is not balanced.

**Proof.** Let $\Gamma j = \{i, k, s\}$. Suppose that $G$ is balanced. Then it follows from Theorem 3.1(a) that $\theta_{ik} - \theta_{sk} = 0$. Similarly $\theta_{is} = \theta_{ik}$, and so, by Lemma 3.12,

$$\theta_{is} = \theta_{ik} = \theta_{sk} = p^2(\theta_e - 1) + p.$$

Hence Theorem 3.1(b) takes the simple form

$$3\theta_e - 1 = p^2(\theta_e - 1) + p,$$

from which the proposition follows at once. □

**Remark 3.17.** Let $H$ be a balanced graph satisfying the conditions of Proposition 3.14 (case 1) or Proposition 3.16 (case 2). Suppose that $x$ and $y$ are two equivalent vertices of $H$, which are either equal to an end point of any of the segments concerned, or do not at all belong to any of those segments. Suppose furthermore that in case 2, $x$ and $y$ are not equal to $j$. Then $\langle x, y, H\rangle \to G$ is balanced only if $G$ is a tree. This can be shown by counting vertices and edges.

The above results (and those of the appendix) have led us to believe that *all* balanced graphs can be obtained through tree-wise joins of substitution results of $k$-fold linear graphs into edge-equivalent graphs.

Let us put this differently and more exactly. We have already shown that a graph with a cut point is balanced if and only if it is the tree-wise join of balanced graphs with a common value of $\theta_e$. The characterization of balanced graphs could be completed now by proving the following conjecture.

**Conjecture 3.18.** *A block is balanced if and only if it can be obtained by substitution of a linear graph into an edge-equivalent graph.*

## Appendix

In Table A.1 we give a survey with respect to balancedness of graphs on 6 or fewer points. Our starting point was [2, Appendix 1].

We have omitted all disconnected graphs from Harary's list, and also all graphs with a bridge. The remaining graphs, listed below, are characterized by three parameters: $n$ (number of points), $r$ (number of edges), and $H$ (the number which Harary in his table assigns to the graphs).

Our criteria have been applied in the same order as in the list of abbreviations below the table.

The eight graphs for which a different argument was necessary are shown in Fig. A.1. The graph (6, 8, 23) is not balanced on account of Proposition 3.16 with $p = 2$. The graph (6, 14, -) is not balanced on account of the remark immediately following Proposition 3.8. The graphs

**Table A.1**
Balancedness of graphs on 6 or fewer points

| *n* | *r* | *H* | |
|---|---|---|---|
| 3 | 3 | – | BE |
| 4 | 4 | 2 | BE |
| 4 | 5 | – | N2 |
| 4 | 6 | – | BE |
| 5 | 5 | 6 | BE |
| 5 | 6 | 1 | BT |
| | | 4 | NV |
| | | 5 | BE |
| 5 | 7 | 1 | N2 |
| | | 2 | N2 |
| | | 4 | N2 |
| 5 | 8 | 1 | N2 |
| | | 2 | NA |
| 5 | 9 | – | NV |
| 5 | 10 | – | BE |
| 8 | 6 | 7 | BE |
| 6 | 7 | 5 | NS |
| | | 6 | NV |
| | | 7 | NS |
| | | 13 | NT |
| 6 | 8 | 1 | NT |
| | | 5 | NV |
| | | 6 | NV |
| | | 7 | NV |
| | | 9 | NV |

| *n* | *r* | *H* | |
|---|---|---|---|
| 6 | 8 | 12 | NV |
| | | 14 | NT |
| | | 15 | BE |
| | | 16 | NV |
| | | 21 | NS |
| | | 23 | NA |
| 6 | 9 | 1 | N2 |
| | | 2 | N2 |
| | | 5 | N2 |
| | | 6 | NT |
| | | 7 | NV |
| | | 8 | N2 |
| | | 9 | N2 |
| | | 10 | N2 |
| | | 11 | N2 |
| | | 13 | N2 |
| | | 16 | N2 |
| | | 17 | BE |
| | | 18 | N2 |
| | | 19 | N2 |
| | | 20 | N2 |
| 6 | 10 | 1 | N2 |
| | | 2 | NA |
| | | 3 | N2 |
| | | 4 | N2 |

| *n* | *r* | *H* | |
|---|---|---|---|
| 6 | 10 | 5 | N2 |
| | | 7 | NA |
| | | 8 | N2 |
| | | 10 | N2 |
| | | 11 | N2 |
| | | 12 | NA |
| | | 14 | NA |
| | | 15 | N2 |
| 6 | 11 | 1 | N2 |
| | | 2 | NV |
| | | 3 | NV |
| | | 5 | N2 |
| | | 6 | NV |
| | | 7 | NV |
| | | 8 | NV |
| | | 9 | N2 |
| 6 | 12 | 1 | NV |
| | | 2 | NV |
| | | 3 | N2 |
| | | 4 | NV |
| | | 5 | BE |
| 6 | 13 | 1 | NV |
| | | 2 | NA |
| 6 | 14 | – | NA |
| 6 | 15 | – | BE |

BE: edge-equivalent, hence balanced.
BT: tree-wise join of edge-equivalent graphs with a common value of $\theta$, hence balanced.
NT: tree-wise join other than above, hence not balanced.
N2: not balanced on account of Proposition 3.8 with $v_0 = 2$.
NV: not balanced on account of Proposition 3.16 with $p = 1$.
NS: not balanced on account of Proposition 3.14.
NA: not balanced on account of a different argument.

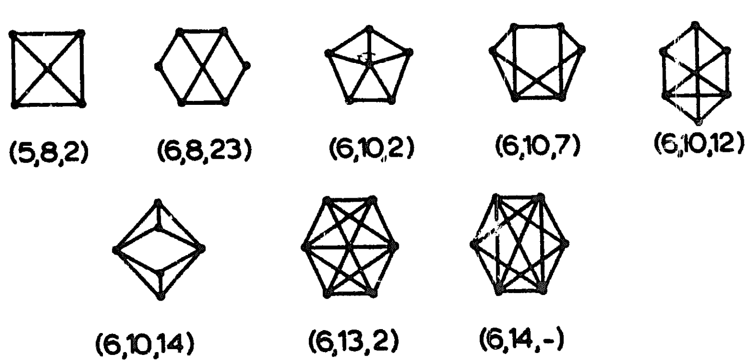


Fig. A.1. Graphs which require special proofs for their unbalancedness.

(6, 10, 12) and (6, 13, 2) satisfy the conditions of Proposition 1.4, while the vertices 1 and 2 are equivalent, so that $\gamma_{12}$ can be quickly found. The remaining four graphs have been dealt with by direct computation.

## Note added in proof

Recently our Conjecture 3.18 has been disproved. A computer search has shown that there exists exactly one balanced graph $G$ on 7 points which is a block but which cannot be obtained by substitution of a linear graph into an edge-equivalent graph. This graph is $G = (X, \Gamma)$ with

$$X = \{1, 2, ..., 7\}, \qquad \Gamma = \{12, 13, 14, 15, 26, 36, 47, 57, 67\}.$$

## References

[1] K.L. Chung, Markov Chains with Stationary Transition Probabilities, 2nd ed. (Springer, Berlin, 1967).

[2] F. Harary, Graph Theory (Addison–Wesley, Reading, Mass., 1969).

[3] J.G. Kemenv and J.L. Snell, Finite Markov Chains (Van Nostrand, New York, reprinted 1969).