Kemeny's constant for several families of graphs and real-world networks

Kooij, Robert E.; Dubbeldam, Johan L.A.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Kemeny's constant for several families of graphs and real-world networks

Robert E. Kooij [*], Johan L.A. Dubbeldam

Faculty of Electrical Engineering, Mathematics and Computer Science, University of Technology Delft, The Netherlands

ARTICLE INFO

ABSTRACT

The linear relation between Kemeny's constant, a graph metric directly linked with random walks, and the effective graph resistance in a regular graph has been an incentive to calculate Kemeny's constant for various networks. In this paper we consider complete bipartite graphs, (generalized) windmill graphs and tree networks with large diameter and give exact expressions of Kemeny's constant. For non-regular graphs we propose two approximations for Kemeny's constant by adding to the effective graph resistance term a linear term related to the degree heterogeneity in the graph. These approximations are exact for complete bipartite graphs, but show some discrepancies for generalized windmill and tree graphs. However, we show that a recently obtained upper-bound for Kemeny's constant in Wang et al. (2017) based on the pseudo inverse Laplacian gives the exact value of Kemeny's constant for generalized windmill graphs. Finally, we have evaluated Kemeny's constant, its two approximations and its upper bound, for 243 real-world networks. This evaluation reveals that the upper bound is tight, with average relative error of only 0.73%. In most cases the upper bound clearly outperforms the other two approximations.

1. Introduction

Kemeny's constant, a graph metric first proposed in 1960 [7], links random walks, Markov chains and spectral graph theory, see for instance [12,14] and [19]. It has already been established that there are several equivalent ways to express Kemeny's constant: using effective graph resistance, random walks, spectral graph theory, pseudo inverse Laplacians, see [6]. An extension of Kemeny's constant to weighted networks (weighted Kemeny's constant) has recently found applications in robotics surveillance [16].

In this paper we consider undirected graphs $G(N, L)$ with N nodes and L links. The adjacency matrix A of a graph G is an $N \times N$ symmetric matrix with elements a_{ij} that are either 1 or 0 depending on whether there is a link between nodes i and j or not. The Laplacian matrix Q of G is an $N \times N$ symmetric matrix $Q = \Delta - A$, where $\Delta = \mathrm{diag}(d_i)$ is the $N \times N$ diagonal degree matrix with the elements $d_i = \sum_{j=1}^{N} a_{ij}$.

A random walk on the graph G gives rise to a Markov chain, with transition probability matrix P satisfying $P = \Delta^{-1}A$. The transition probability matrix P of a finite, irreducible Markov chain, and its steady state probability vector π and the all-ones vector u, satisfy $Pu = u$ and $\pi^T P = \pi^T$.

---

* Corresponding author.
  E-mail address: r.e.kooij@tudelft.nl (R.E. Kooij).

Kemeny defined his constant, in terms of the matrix Z, which, for any two column vectors h and g such that the scalar products $h^T u$ and $\pi^T g$ are nonzero, is given by

$$Z \equiv \left(I - P + gh^T\right)^{-1}. \tag{1}$$

The Kemeny constant is defined, in terms of the trace of the matrix Z, as

$$K(P) \equiv \text{trace}\,(Z) - \pi^T Z u. \tag{2}$$

For a given transition probability matrix P and with $h^T g = 1$, the Kemeny constant $K(P)$ is the same regardless of the choice of the matrix Z defined above.

A direct relation between Kemeny's constant and random walks was given by Kemeny and Snell [7]. In fact, they state that the expression $\tilde{K}$ defined as

$$\tilde{K}(P) = \sum_{i=1}^{N} \pi_i m_{ji}, \tag{3}$$

where π_i denotes the stationary probability for node i and m_{ji} denotes the mean first passage time (the expected number of steps before node i is visited, starting from node j), is a constant and hence independent of the starting position j.

The relation between $\tilde{K}$ and K is given by

$$\tilde{K}(P) = K(P) + 1. \tag{4}$$

Combining Eqs. (3) and (4) we get

$$K(P) = \sum_{i=1}^{N} \pi_i m_{ji} - 1. \tag{5}$$

It is a bit confusing that there are actually two definitions for Kemeny's constant. This is due to the fact that for ergodic Markov chains $\{X_n\}$, there are two different but related random variables involving time. Hitting time is defined as $T_i = min\{n \geq 0 : X_n = i\}$, while the recurrence time is given by $T_i^+ = min\{n \geq 1 : X_n = i\}$. Clearly, $E[T_i] = 0$ while $E[T_i^+] \geq 1$. It is well-known, see [12], that $E[T_i^+] = \frac{1}{\pi_i}$. Therefore $\pi_i E[T_i] = 0$ whereas $\pi_i E[T_i^+] = 1$, while both quantities may be equivocally be denoted as $\pi_i m_{ii}$. In this paper we will follow the definition as given by Eq. (2) for Kemeny's constant.

Since Eq. (3) is independent of j, by summing over all nodes, the "random surfer" interpretation of $\tilde{K}$ by Levene and Loizou [11] follows:

$$\tilde{K}(P) = \sum_{i=1}^{N} \pi_i \frac{\sum_{j=1}^{N} m_{ji}}{N}. \tag{6}$$

Lovasz [12] showed that the right-hand side of (5) can be expressed in terms of the eigenvalues $\lambda_1, \lambda_2 \ldots \lambda_N$ of the symmetric matrix S, defined as $S = \Delta^{-1/2} A \Delta^{-1/2}$, where $\lambda_1 = 1 > \lambda_2 \geq \lambda_2 \ldots \lambda_N \geq -1$:

$$K(P) = \sum_{k=2}^{N} \frac{1}{1 - \lambda_k}. \tag{7}$$

The relation between Kemeny's constant and the effective graph resistance has been made explicit for regular graphs by Palacios et al. [13]. Recently, Wang et al. [19] presented the following closed-form formula for Kemeny's constant, in terms of the Moore–Penrose pseudo inverse $Q^\dagger$ of the Laplacian matrix,

$$K(P) = \zeta^T d - \frac{d^T Q^\dagger d}{2L}, \tag{8}$$

where the column vector $\zeta = \left(Q_{11}^\dagger,\ Q_{22}^\dagger,\ \ldots,\ Q_{NN}^\dagger\right)$ and $d = (d_1,\ d_2,\ \ldots,\ d_N)$ denotes the degree vector for the graph.

In this paper we will determine Kemeny's constant for a number of families of graphs and real-world networks, by using either Eqs. (5), (7) or (8).

The paper is organized as follows. Section 2 presents analytic expressions for Kemeny's constant for three graph families. The relation between Kemeny's constant and the effective graph resistance if further explored in Section 3. In Section 4 we evaluate an upper bound for Kemeny's constant. Section 5 concludes the paper.

2. Graph families

In this section we derive analytic expressions for Kemeny's constant for three different graph families.

2.1. Complete bipartite graphs

We start with Kemeny's constant for complete bipartite graphs, using Eq. (5). A complete bipartite graph K_{N_1,N_2} consists of two disjoint sets S_1 and S_2 containing, respectively, N_1 and N_2 nodes, such that all nodes in S_1 are connected to all nodes in S_2, while within each set no connections occur.

Theorem 1. *Kemeny's constant for the complete bipartite graph K_{N_1,N_2} is given by:*

$$K(P) = N_1 + N_2 - \frac{3}{2}. \tag{9}$$

Proof. We can prove the theorem from the definition of the Kemeny constant given in Eq. (5). For the complete bipartite graph K_{N_1,N_2} the matrix $P = \Delta^{-1}A$ reads

$$P = \begin{pmatrix} 0_{N1 \times N1} & \frac{1}{N_1}J_{N1 \times N2} \\ \frac{1}{N_2}J_{N2 \times N1} & 0_{N2 \times N2} \end{pmatrix} \tag{10}$$

where J denotes the all-ones matrix. The left-eigenvector of P with eigenvalue 1 is easily seen to be
$\pi^T = \left(\frac{1}{2N_1} \cdots \frac{1}{2N_1}; \frac{1}{2N_2} \cdots \frac{1}{2N_2} \right)$, which contains N_1 times $\frac{1}{2N_1}$ and N_2 times $\frac{1}{2N_2}$

We can now apply the definition given by Eq. (5). As the Kemeny constant does not depend on j, we may set $j = 1$, which we assume corresponds to a node in the set S_1. Then we find

$$K(P) = \sum_{i=1}^{N_1+N_2} \pi_i m_{1i} - 1 = \frac{1}{2} m_{1S_1} + \frac{1}{2} m_{1S_2} - 1, \tag{11}$$

where m_{1S_1} denotes the mean passage time from node 1 (which belongs to S_1) to a specific node in S_1. Likewise, m_{1S_2} denotes the mean passage time from node 1 to a specific node in S_2. Then, conditioning on the first jump, we obtain

$$m_{1S_2} = \frac{1}{N_2} + (1 - \frac{1}{N_2})(2 + m_{1S_2}), \tag{12}$$

which leads to

$$m_{1S_2} = 2N_2 - 1, \tag{13}$$

In a similar way we obtain

$$m_{S_21} = 2N_1 - 1, \tag{14}$$

where m_{S_21} denotes the mean passage time from any node in S_2 to node 1 in S_1. Finally,

$$m_{1S_1} = 1 + m_{S_21} = 1 + 2N_1 - 1 = 2N_1. \tag{15}$$

Combining Eqs. (11), (13) and (15) gives

$$K(P) = N_1 + N_2 - \frac{3}{2}. \quad \square$$

2.2. Trees with a large diameter

In recent work [8] Kirkland and Zeng have derived a general expression for Kemeny's constant on trees in terms of the degree sequence and distance matrix, in order to demonstrate that Kemeny's constant will increase upon inserting an edge between so-called twin pendent vertices. The expression for K in [8] was only made explicit, in terms of the number of nodes, for the path graph P_N. Therefore, we give such explicit expressions for K for a number of trees with large diameter here.

The path graph P_N is obviously a tree with N nodes and diameter $N - 1$. To determine Kemeny's constant for P_N we use Eq. (5). We will need the following well-known facts, see [12], about the stationary distribution π and the average return time m_{ii}:

$$\pi_i = \frac{d_i}{2L}, \quad m_{ii} = \frac{2L}{d_i}. \tag{16}$$

An expression for the mean hitting times for the path graph is also given in [12]: for $1 \le i < k \le N$ it satisfies:

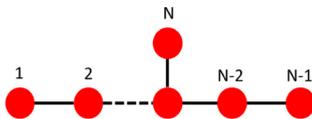$$m_{ik} = (k-1)^2 - (i-1)^2. \tag{17}$$

Fig. 1. The graph D_N.

Fig. 2. The graph $E_{N,2}$.

Theorem 2. *Kemeny's constant for the path graph P_N is given by:*

$$K(P) = \frac{1}{3}N^2 - \frac{2}{3}N + \frac{1}{2}. \tag{18}$$

Proof. We apply Eq. (5) and take $j = 1$. Then, we have $K(P) = \sum_{i=1}^{N} \pi_i m_{1i} - 1 = \pi_1 m_{11} + \sum_{i=2}^{N-1} \pi_i m_{1i} + \pi_N m_{1N} - 1$.
Applying, Eqs. (16) and (17) this leads to $K(P) = \frac{1}{N-1} \sum_{i=2}^{N-1} (i-1)^2 + \frac{1}{2}(N-1)$. Finally, using the identity

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}, \tag{19}$$

leads to Eq. (18). □

Next we consider trees with diameter $N - 2$. First we look at trees which are sometimes denoted as D_N, see [2]. The graph consists of a path of length $N - 2$, with the nodes labeled, from left to right, from 1 to $N - 1$. Node N is connected to node $N - 2$, see Fig. 1.

Theorem 3. *Kemeny's constant for the tree D_N is given by:*

$$K(P) = \frac{\frac{2}{3}N^3 - 2N^2 - \frac{5}{3}N + 11}{2(N-1)}. \tag{20}$$

Proof. Again we will apply Eq. (5) with $j = 1$, leading to $K(P) = \sum_{i=2}^{N-3} \pi_i m_{1i} + \pi_{N-2} m_{1,N-2} + \pi_{N-1} m_{1,N-1} + \pi_N m_{1N}$.
Applying, Eqs. (16) and (17), and the observations that $\pi_{N=1} = \pi_N = \frac{1}{2N-1}$ and $m_{1,N-1} = m_{1N}$, leads to $K(P) = \frac{1}{N-1} \sum_{i=2}^{N-3} (i-1)^2 + 3 \frac{(N-3)^2}{2(N-1)} + \frac{1}{N-1} m_{1N}$.
It is easy to see that $m_{1N} = m_{1,N-2} + m_{N-2,N}$. From Eq. (16) we have $m_{1,N-2} = (N-3)^2$. Because $m_{N-2,N}$ satisfies $m_{N,N} = 1 + m_{N-2,N}$, where $m_{N,N} = 2(N-1)$ according to Eq. (16), we have $m_{N-2,N} = 2N - 3$. Using the above identities and Eq. (19) leads to Eq. (20). □

Next we look at a broader class of trees with diameter $N - 2$, which we will denote as $E_{N,M}$. The graph consists of a path of length $N - 2$, with the nodes labeled, from left to right, from 1 to $N - 1$. Node N is connected to node $N - M - 1$. For example, Fig. 2 displays the graph $E_{N,2}$.

The path P_N and the tree D_N are special cases of $E_{N,M}$, as $P_N = E_{N,0}$ and $D_N = E_{N,1}$.

Theorem 4. *Kemeny's constant for the tree $E_{N,M}$ is given by:*

$$K(P) = \frac{1}{3}N^2 - \frac{2}{3}N + \frac{1}{2} - \frac{2(M(N-3) - M(M-1))}{N-1}. \tag{21}$$

The proof of Theorem 4 is basically the same as the proof of Theorem 3, although a bit more elaborate. We leave it to the reader as an exercise.

2.3. Generalized windmill graphs

In this subsection we will consider a generalization of the class of so-called windmill graphs. A windmill graph $W(\eta, k)$ consists of η copies of the complete graph K_k, with every node connected to a common node, see Fig. 3.

Recently Estrada [4] studied windmill graphs and showed that the clustering coefficient and the transitivity index of such graphs diverge, when the graph size tends to infinity. It is shown in [4] that windmill graphs are better suited to model certain real-world networks, then classical network models such as Erdős–Rényi and Barabási–Albert (BA)
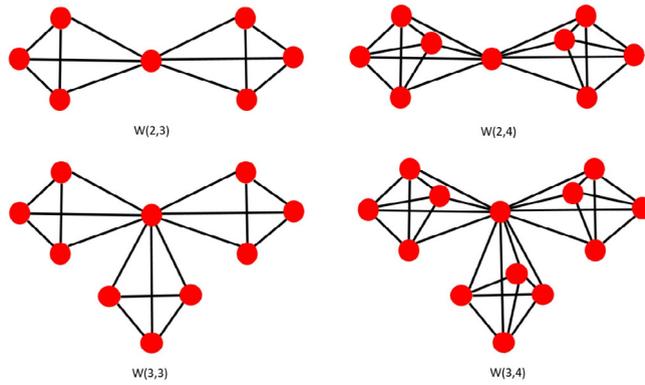
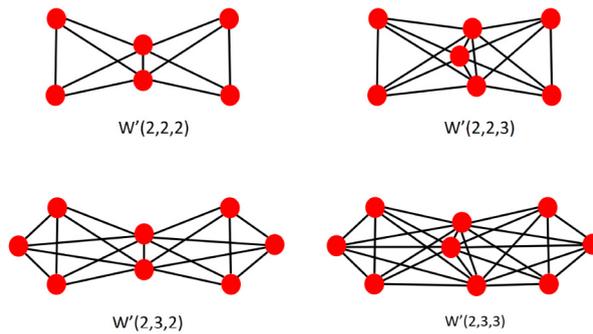**Fig. 3.** Illustration of some windmill graphs $W(\eta, k)$.

Fig. 4. Some generalized windmill graphs $W'(\eta, k, l)$ of Type I.
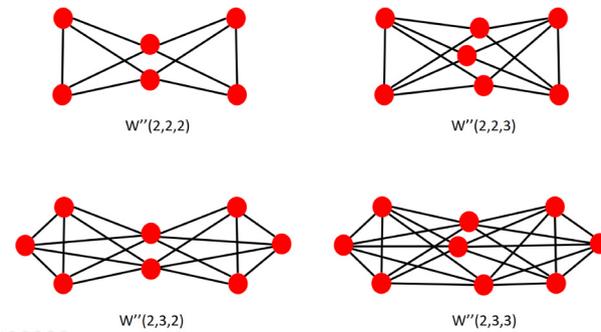

Fig. 5. Some generalized windmill graphs $W''(\eta, k, l)$ of Type II.

networks. Estrada [4] also studied the spectra of the adjacency and the Laplacian matrices of these graphs. In this paper we will determine Kemeny's constant both for the windmill graphs and for two generalizations of these graphs, recently suggested by Kooij [10]. Generalized windmill graphs can be used to described public transportation networks in so-called P-space [10]. For both generalizations we replace the central node, connecting all η copies of the complete graph K_k, by l central nodes. For the first generalization, we assume the l central nodes are all connected, i.e. they form a clique K_l. We call this a generalized windmill graph of Type I and denote it by $W'(\eta, k, l)$. Obviously, it holds that $W'(\eta, k, 1) = W(\eta, k)$. For the second generalization, we assume the l central nodes have no connections among each other. We will refer to it as a generalized windmill graph of Type II and denote it by $W''(\eta, k, l)$. Figs. 4 and 5 depict examples of the generalized windmill graphs of Type I and II, respectively. Note that the generalized windmill graph of Type I was introduced recently also independently by Estrada and Benzi [5], who refer to it as a core-satellite graph.

We will now determine Kemeny's constant for the generalized windmill graph $W'(\eta, k, l)$ of Type I. As a starting point we will give the adjacency matrix for $W'(\eta, k, l)$ and denote it by $A'(W')$.

$$A'(W'(\eta, k, l)) = \begin{bmatrix} (J-I)_{l \times l} & J_{l \times k} & J_{l \times k} & \dots & J_{l \times k} \\ J_{k \times 1} & (J-I)_{k \times k} & 0_{k \times k} & \dots & 0_{k \times k} \\ J_{k \times 1} & 0_{k \times k} & (J-I)_{k \times k} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ J_{k \times 1} & 0_{k \times k} & \dots & \dots & (J-I)_{k \times k} \end{bmatrix}. \tag{22}$$

The degree distribution for $W'(\eta, k, l)$ is bi-modal: the l core nodes all have degree $d_{core} = \eta k + l - 1$, while the ηk nodes in the η cliques have degree $d_{clique} = k - 1 + l$.

From this it follows that the degree matrix Δ' satisfies

$$\Delta' = \begin{bmatrix} d_{core}I_{l \times l} & 0_{l \times \eta k} \\ 0_{\eta k \times l} & d_{clique}I_{\eta k \times \eta k} \end{bmatrix}. \tag{23}$$

Combining this with (22), we obtain the expression for the symmetric matrix S':

$$S' = \begin{bmatrix} r(J-I)_{l \times l} & sJ_{l \times k} & sJ_{l \times k} & \dots & sJ_{l \times k} \\ sJ_{k \times 1} & t(J-I)_{k \times k} & 0_{k \times k} & \dots & 0_{k \times k} \\ sJ_{k \times 1} & 0_{k \times k} & t(J-I)_{k \times k} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ sJ_{k \times 1} & 0_{k \times k} & \dots & \dots & t(J-I)_{k \times k} \end{bmatrix}. \tag{24}$$

with $r = \frac{1}{\eta k + l - 1}$, $s = \frac{1}{\sqrt{\eta k + l - 1}\sqrt{k - 1 + l}}$ and $t = \frac{1}{k - 1 + l}$.

Lemma 5. *The spectrum of the matrix $S' = (\Delta')^{-1/2}A'(\Delta')^{-1/2}$, for the generalized windmill graph $W'(\eta, k, l)$ of Type I, is*

$$\{(-\frac{1}{k + l - 1})^{\eta(k-1)}, (\frac{k - 1}{k + l - 1})^{\eta - 1}, (-\frac{1}{\eta k + l - 1})^{l-1},$$
$$(\frac{l - 1}{\eta k + l - 1} - \frac{l}{k + l - 1})^1, (1)^1\} \tag{25}$$

The proof of Lemma 5 is given in Appendix A.

We are now in the position to determine Kemeny's constant for the generalized windmill graph of Type I.

Theorem 6. *Kemeny's constant for the generalized windmill graph of Type I $W'(\eta, k, l)$ is given by:*

$$K(P) = \frac{\eta(k-1)(k+l-1)}{k+l} + \frac{(\eta-1)(k+l-1)}{l} + \frac{(l-1)(\eta k+l-1)}{\eta k+l}$$
$$+ \frac{(\eta k+l-1)(k+l-1)}{\eta k(k+l-1)+l(\eta k+l-1)}. \tag{26}$$

Proof. The theorem follows directly from plugging the eigenvalues given in Eq. (25) into Eq. (7). □

Corollary 1. *Kemeny's constant for the windmill graph $W(\eta, k)$ is given by:*

$$K(P) = \frac{k^2(2\eta - 1)}{k + 1}. \tag{27}$$

Proof. Because $W(\eta, k)$ corresponds to $W'(\eta, k, l)$ with $l = 1$, the result follows from the substitution of $l = 1$ into Eq. (26). □

We will now derive Kemeny's constant for the generalized windmill graph of Type II, $W''(\eta, k, l)$. As a first step we will give the form of the symmetric matrix $S'' = (\Delta'')^{-1/2}A''(\Delta'')^{-1/2}$, where A'' denotes the adjacency matrix of the generalized windmill graph of Type II $W''(\eta, k, l)$. Analogous to the case of the generalized windmill graph of Type I, we can show that

$$S'' = \begin{bmatrix} 0_{l \times l} & s'J_{l \times k} & s'J_{l \times k} & \dots & s'J_{l \times k} \\ s'J_{k \times 1} & t'(J-I)_{k \times k} & 0_{k \times k} & \dots & 0_{k \times k} \\ s'J_{k \times 1} & 0_{k \times k} & t'(J-I)_{k \times k} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ s'J_{k \times 1} & 0_{k \times k} & \dots & \dots & t'(J-I)_{k \times k} \end{bmatrix}, \tag{28}$$

with $s' = \frac{1}{\sqrt{\eta k}\sqrt{k - 1 + l}}$ and $t' = \frac{1}{k - 1 + l}$.

Lemma 7. *The spectrum of the matrix $S'' = (\Delta'')^{-1/2}A''(\Delta'')^{-1/2}$, for the generalized windmill graph $W''(\eta, k, l)$ of Type II, is*

$$\{(-\frac{1}{k+l-1})^{\eta(k-1)}, (\frac{k-1}{k+l-1})^{\eta-1}, (0)^{l-1}, (-\frac{l}{k+l-1})^1, (1)^1\} \tag{29}$$

The proof of Lemma 7 is given in Appendix B.

Again Kemeny's constant for the generalized windmill graph of Type II follows from application of Eq. (7).

Theorem 8. *Kemeny's constant for the generalized windmill graph of Type II $W''(\eta, k, l)$ is given by:*

$$K(P) = \frac{\eta(k-1)(k+l-1)}{k+l} + \frac{(\eta-1)(k+l-1)}{l} + l - 1 + \frac{k+l-1}{k+2l-1}. \tag{30}$$

Proof. The theorem follows directly from plugging the eigenvalues given in Eq. (29) into Eq. (7). $\square$

3. Relation with effective graph resistance

For a regular graph on N nodes with degree r, the relation between Kemeny's constant and the effective graph resistance was shown [15] to be

$$K(P) = \frac{r}{N}R_G, \tag{31}$$

where R_G denotes the effective graph resistance. In this section we will propose two approximations for Kemeny's constant for non-regular graphs, inspired by Eq. (31).

3.1. A first approximation for kemeny's constant

Before we give our first approximation, we introduce some notation. For a graph on N nodes and L links we denote by D the average degree of the nodes, i.e. $D = \frac{2L}{N}$. The heterogeneity index H, a metric which quantifies the variability of the degree distribution, see [1], is defined as follows:

$$H = \frac{1}{N}\sum_{i=1}^{N}(d_i - D)^2, \tag{32}$$

where d_i denotes the degree of node i. Now, we assume that the approximation for Kemeny's constant takes the following form:

$$K^*(P) = \frac{D}{N}R_G + Hf(N, L), \tag{33}$$

where $f(N, L)$ is a function that still needs to be determined. For regular graphs with degree r, Eq. (33) simplifies to Eq. (31) because for that case $D = r$ and $H = 0$.

We will determine $f(N, L)$ by considering the case of complete bipartite graphs K_{N_1, N_2}, discussed in Section 2.1. For K_{N_1, N_2} we have $N = N_1 + N_2$ and $L = N_1 N_2$. Therefore

$$D = \frac{2N_1 N_2}{N_1 + N_2}. \tag{34}$$

An elementary calculation further shows that

$$H = \frac{N_1 N_2(N_1 - N_2)^2}{(N_1 + N_2)^2}. \tag{35}$$

The Laplacian spectrum for K_{N_1, N_2}, satisfies $\{0^1, N_1^{N_2-1}, N_2^{N_1-1}, (N_1 + N_2)^1\}$, see [17]. Hence the effective graph resistance for K_{N_1, N_2} satisfies:

$$R_G = (N_1 + N_2)(\frac{N_2 - 1}{N_1} + \frac{N_1 - 1}{N_2} + \frac{1}{N_1 + N_2}). \tag{36}$$

Plugging the result of Theorem 1 and the above expressions into Eq. (33) we get

$$N_1 + N_2 - \frac{3}{2} = \frac{2N_1 N_2}{N_1 + N_2}(\frac{N_2 - 1}{N_1} + \frac{N_1 - 1}{N_2} + \frac{1}{N_1 + N_2}) + \frac{N_1 N_2(N_1 - N_2)^2}{(N_1 + N_2)^2}f. \tag{37}$$

After some manipulation of Eq. (37) we obtain an explicit expression for f:

$$f = \frac{1 - 2N_1 - 2N_2}{2N_1 N_2}. \tag{38}$$

Given that for K_{N_1,N_2} we have $N = N_1 + N_2$ and $L = N_1 N_2$, we assume the following expression for $f(N, L)$ for general graphs:

$$f(N, L) = \frac{1 - 2N}{2L}. \tag{39}$$

Hence, we have established the following result.

Theorem 9. *For a graph on N nodes and L links, denote the heterogeneity index and effective graph resistance by H and R_G, respectively. Let*

$$K^*(P) = \frac{2L}{N^2}R_G + H\frac{1 - 2N}{2L}. \tag{40}$$

Then, for complete bipartite graphs, $K^*(P)$ is equal to Kemeny's constant.

3.2. Applying the approximation $K^*(P)$ to windmill graphs

In this section we will show that the approximation K^* is also exact for windmill graphs i.e. graphs $W(\eta, k)$. According to [4] for $W(\eta, k)$ we have $N = \eta k + 1$ and $L = \frac{\eta k(k+1)}{2}$. Therefore

$$D = \frac{\eta k(k+1)}{\eta k + 1}. \tag{41}$$

An elementary calculation further shows that

$$H = \frac{\eta(\eta - 1)^2 k^3}{(\eta k + 1)^2}. \tag{42}$$

The Laplacian spectrum for $W(\eta, k)$ satisfies $\{0^1, 1^{\eta-1}, (k+1)^{\eta(k-1)}, (\eta k + 1)^1\}$, see [4]. Hence the effective graph resistance for $W(\eta, k)$ satisfies:

$$R_G = (\eta k + 1)(\eta - 1 + \frac{\eta(k-1)}{k+1} + \frac{1}{\eta k + 1}). \tag{43}$$

Now, using the expressions for N, L, H and R_G, we can show that for $W(\eta, k)$ Eq. (40) gives

$$K^*(P) = \frac{(2\eta - 1)k^2}{k+1}. \tag{44}$$

Hence we have also proved the following result.

Theorem 10. *For a graph on N nodes and L links, denote the heterogeneity index and effective graph resistance by H and R_G, respectively. Let*

$$K^*(P) = \frac{2L}{N^2}R_G + H\frac{1 - 2N}{2L}. \tag{45}$$

Then, for the windmill graphs $W(\eta, k)$, K^* is equal to Kemeny's constant.

We have chosen to use the heterogeneity index H to quantify the heterogeneity of the degree distribution. However, there are several other heterogeneity metrics, see [3]. In the next subsection we derive another approximation for $K(P)$ based upon a variant of the so-called irregularity index.

3.3. A second approximation for Kemeny's constant

An alternative to the heterogeneity index H, is a variant of the irregularity index, see [18], defined as

$$I = \lambda_1^2 - D^2, \tag{46}$$

where λ_1 denotes the largest eigenvalue of the adjacency matrix. Note that the original index was defined as $\lambda_1 - D$ and that $I = 0$ for regular graphs. Then, in a way similar to the proof of Theorem 9, we can obtain the following result:

Theorem 11. *For a graph on N nodes and L links, denote the irregularity index and effective graph resistance by I and R_G, respectively. Let*

$$K^{**}(P) = \frac{2L}{N^2}R_G + I\frac{1 - 2N}{2L}. \tag{47}$$

Then, for complete bipartite graphs, $K^{**}(P)$ is equal to Kemeny's constant.

Table 1
Kemeny's constant and its approximations K^* and K^{**} for several graphs.

Graph	N	L	$K(P)$	$K^*(P)$	$K^{**}(P)$
$K_{10,15}$	25	150	23.50	23.50	23.50
P_{10}	10	9	27.17	29.53	29.23
D_{10}	10	9	25.61	28.06	27.27
$E_{10,2}$	10	9	24.50	27.16	26.71
$W(3,10)$	31	165	45.45	45.45	43.88
$W'(3,10,5)$	35	295	35.49	33.77	30.51
$W''(3,10,5)$	35	285	35.54	34.67	33.30

Table 2
Kemeny's constant and the upper bound $K_U(P)$, for several graphs.

Graph	$K(P)$	$K_U(P)$
$K_{10,15}$	23.50	23.50
P_{10}	27.17	27.28
D_{10}	25.61	25.71
$E_{10,2}$	24.50	24.61
$W(3,10)$	45.45	45.45
$W'(3,10,5)$	35.49	35.49
$W''(3,10,5)$	35.54	35.54

3.4. Evaluation of the approximations for other graphs

In the previous two subsections we have shown that the approximations K^* and K^{**} equal Kemeny's constant for the cases of complete bipartite graphs, while K^* also equals Kemeny's constant for windmill graphs.

In this subsection we will show that in general K^* and K^{**} do not equal Kemeny's constant, by comparing K^*, K^{**} and K, for some realizations of graphs we have studied in Section 2. Table 1 shows the results of this comparison.

Table 1 illustrates that for the considered trees and generalized windmill graphs, K^* and K^{**} do not equal Kemeny's constant. For the trees both approximations overestimate Kemeny's constant while for the generalized windmill graphs they give an underestimation.

4. An upper bound for kemeny's constant

In [19] not only Eq. (8) was derived but also a closely connected upper bound:

$$K(P) \le \zeta^T d - \frac{H}{D\mu_1} \equiv K_U(P), \tag{48}$$

where μ_1 denotes the largest Laplacian eigenvalue. In this section we will evaluate how tight the upper bound $K_U(P)$ is. We will again consider the graph models considered in the previous sections.

Table 2 gives the considered graphs, Kemeny's constant and the upper bound K_U.

Table 2 shows that, for the considered graphs, the upper bound $K_U(P)$ is rather tight. The largest relative error for $K_U(P)$ in Table 2 is only 0.4%, which occurs for $E_{10,2}$.

For both the windmill and generalized windmill graphs, Table 2 contains numerical evidence that for these graph families, $K_U(P)$ actually equals Kemeny's constant. The proof this is indeed the case for any generalized windmill graph hinges on the fact that all eigenvectors of the Laplacian matrix Q, except the one corresponding to the largest eigenvalue, are orthogonal to the degree vector d of the graph.

Proposition 1. *For the graphs $W(\eta, k)$, $W'(\eta, k, l)$ and $W''(\eta, k, l)$ the upper bound given in Eq. (48) is tight.*

Proof. We will proof the statement that for generalized windmill graphs the Kemeny constant reduces to $\zeta^T d - \frac{H}{D\mu_1}$. For that we need to calculate the pseudo inverse Laplacian $Q^\dagger$, which can be easily done once the eigenvalues and eigenvectors of the Laplacian matrix are known. The Laplacian spectra for generalized windmill graphs are given by Kooij [10] and generalize the results of Estrada [4]. Starting with generalized windmill graphs of Type I (where $l = 1$ reduces to the ordinary windmill graph), the spectrum in this case is

$$Sp(Q(W')) = \{(\eta k + l)^l, (k + l)^{\eta(k-1)}, l^{\eta-1}, 0^1\},$$

where the superscript denotes the multiplicity. The corresponding eigenvectors can be found by inspection. The first normalized eigenvector v_1, with eigenvalue $\eta k + l$, is

$$v_1 = (\eta k, \ldots, \eta k, -l, \ldots, -l)/\sqrt{\eta kl(\eta k + l)} = (\eta k_{1 \times l}, -l_{1 \times \eta k})/\sqrt{\eta kl(\eta k + l)}. \tag{49}$$

The other $(l-1)$ orthogonal eigenvectors belonging to this eigenvalue are of the form

$$(\alpha_1, \alpha_2, \ldots, \alpha_l, 0, \ldots, 0),$$

with $\alpha_1 + \alpha_2 + \cdots + \alpha_l = 0$, and not all α_i equal to zero. The eigenvectors for the other eigenvalues are direct generalizations of the ordinary windmill graphs studied in [4], that is, the all-ones vector u is an eigenvector with eigenvalue 0, and the eigenvalue $k + l$ has eigenvectors of the form $t = (0_{1 \times l}, \alpha_{1_1}, \ldots, \alpha_{1_k}, \alpha_{2_1}, \ldots \alpha_{2_k}, \alpha_{\eta_1} \ldots \alpha_{\eta_k})$, with $\sum_{j=1}^{k} \alpha_{mj} = 0$ for all m and $\alpha_{mj} \neq 0$ for some m and j. Finally, the vector t, with the entries satisfying $\alpha_{mj} = \alpha_m$ for all m and $\alpha_1 + \cdots + \alpha_\eta = 0$ with $\alpha_m \neq 0$ for some m, induces a family of $(\eta - 1)$ eigenvectors with eigenvalue l.

One can easily verify that all eigenvectors are orthogonal to the degree vector d, except the eigenvector v_1. If we next use the representation of $Q^\dagger$ in the eigenbasis, that is

$$Q^\dagger = \sum_{i=1}^{N-1} \frac{1}{\mu_i} v_i v_i^T,$$

we find that in the term $d^T Q^\dagger d$ only a single term survives, which gives

$$d^T Q^\dagger d = \frac{(v_1^T d)^2}{\mu_1}.$$

The last step consists of showing that $(v_1^T d)^2 = HN$. This follows from the fact that $d = Du + \delta$ and $\delta = \sum_{i=1}^{N-1} (v_i^T \delta) v_i$, as the eigenvectors form a orthogonal basis of the complement of u^T, which implies immediately that

$$\delta^T \delta = \sum_{i=1}^{N=1} (v_i^T \delta)^2 = \sum_{i=1}^{N} (d_i - D)^2,$$

see also [19]. For the generalized windmills of Type II, the spectrum is

$$\text{Sp}(Q(W'')) = \{(\eta k + l)^1, (\eta k)^{l-1}, (k + l)^{\eta(k-1)}, l^{\eta-1}, 0^1\},$$

with largest eigenvalue $\eta k + l$, whose eigenspace is spanned by

$$w_1 = (\eta k, \eta k, \ldots, \eta k, -l, \ldots, -l) / \sqrt{\eta k l (\eta k + l)}.$$

The $(l-1)$-dimensional eigenspace corresponding to ηk is spanned by $(\alpha_1, \alpha_2, \ldots, \alpha_l, 0, \ldots, 0)$ with $\alpha_1 + \alpha_2 + \cdots + \alpha_l = 0$. We note again that also in this case all eigenvectors are orthogonal to the degree vector d, except the eigenvector w_1. The proof is therefore analogous to that of the generalized windmill graphs of Type I. For completeness we give the expression for $Q^\dagger$:

$$Q^\dagger = \frac{1}{\mu_1} w_1 w_1^T + \sum_{i=l+1}^{\eta k + l} \frac{1}{\mu_i} w_i w_i^T.$$

Analogous to the Type I case we can prove that $d^T Q^\dagger d$ equals $\sum_{i=1}^{N} (d_i - D)^2 / \mu_1$. $\quad \square$

5. Kemeny's constant, its approximations and upper bound, for real-world networks

So far, we have determined Kemeny's constant $K(P)$, its approximations K^* and K^{**} and the upper bound K_U, for some families of highly structured graphs, such as trees, complete bipartite graphs and (generalized) windmill graphs. In this section we will study Kemeny's constant and its approximations, for a large number of real-world networks. As data source we use the Internet Topology Zoo [9], a collection of more than 250 IP (Internet Protocol) network topologies from around the world. The dataset is available at http://www.topology-zoo.org/. After discarding the networks that are disconnected, we wind up with 243 connected networks.

In Table 3 we show the number of nodes and links of the three smallest and largest networks in our dataset, together with the values of $K(P)$ and its two approximations and upper bound. Kemeny's constant has been determined using Eq. (8).

It can be observed that the approximations and the upper bound are exact for the Renam and Mren networks. This is because both networks are star topologies, and hence this observation follows from Theorems 9 and 11. In a similar fashion we can conclude from Table 3 that Arpanet2012 is not a ring topology, as the approximations are not exact. Finally, we remark that for the large networks, the upper bound K_U is very tight, with relative error below 0.1%, and it clearly outperforms K^* and K^{**}.

Table 4 shows some statistics related to the absolute values of the relative errors for the two approximations and upper bound, evaluated over all 243 real-world networks.

Again we see that K_U is a tight upper bound and in general it outperforms the two approximations K^* and K^{**}. In the list of 243 real-world networks, for 8 networks, the two approximations and the upper bound give the exact value of Kemeny's constant. The reason is that these networks are either regular (Globalcenter, a complete graph and Sanren and Telecomserbia, which are both cycle graphs) or either form a star topology (Basnet, Itnet, Mren, Renam and Singaren).

Table 3
Kemeny's constant, the approximations K^*, K^{**} and the upper bound K_U, for the smallest and largest networks in the Internet Topology Zoo.

Graph	N	L	$K(P)$	K^*	K^{**}	K_U
Arpanet196912	4	4	2.54	2.73	2.55	2.60
Renam	5	4	3.50	3.50	3.50	3.50
Mren	6	5	4.50	4.50	4.50	4.50
UsCarrier	158	189	1175.99	1265.48	1263.39	1176.68
Cogentco	197	245	1082.45	1197.24	1191.59	1083.35
Kdl	754	899	5907.29	6264.78	6261.74	5908.32

Table 4
Statistics for the absolute value of the relative errors for the approximations K^*, K^{**} and the upper bound K_U, for 243 real-world networks.

Metric	K^*	K^{**}	K_U
Average absolute rel. error	27.25%	19.86%	0.73%
Maximum absolute rel. error	122.60%	60.26%	8.05%

For the remaining 235 networks, the following statistics have been found:

- In 230 cases, K_U is the closest to $K(P)$, while in the remaining 5 cases, K^* is the closest to $K(P)$.
- In 3 out of the 235 cases K^* is a better approximation than K^{**}.
- In 1 out of the 235 cases it holds that $K^* < K(P)$, while in 8 out of the 235 cases it holds that $K^{**} < K(P)$.

6. Conclusion

First we have studied Kemeny's constant for several highly structured graphs, including trees with large diameter, (generalized) windmill graphs and complete bipartite graphs. These graphs allow exact evaluation of Kemeny's constant $K(P)$. Using the known relation between the effective graph resistance and Kemeny's constant for regular graphs, we propose two generalizations of this relation by taking into account the presence of degree heterogeneity. The generalization K^*, which depends linearly on the heterogeneity index H, leads to an exact expression for Kemeny's constant, for the case of windmill graphs and complete bipartite graphs. The generalization K^{**}, which depends linearly on a variant of the irregularity index I, leads to an exact expression for Kemeny's constant, for the case of complete bipartite graphs. Next it is proved that an upper bound K_U for Kemeny's constant found by Wang et al. [19], is tight for (generalized) windmill graphs. Finally, we have evaluated Kemeny's constant, its two approximations and its upper bound, for 243 real-world networks. This evaluation reveals that K_U is a tight upper bound, with average relative error of only 0.73%. In most cases K_U clearly outperforms the other two approximations.

CRediT authorship contribution statement

Robert E. Kooij: Conceptualization, Methodology, Software, Writing - original draft. **Johan L.A. Dubbeldam:** Conceptualization, Methodology, Writing - review & editing.

Acknowledgments

Appendix A. Proof of Lemma 5

Let $\mathbf{v}_1 = [0, \ldots, 0, \alpha_{11}, \ldots, \alpha_{1k}, \alpha_{21}, \ldots, \alpha_{2k}, \ldots, \alpha_{\eta 1}, \ldots, \alpha_{\eta k}]^T$, where the first l entries are zero, be a N-dimensional vector such that $\sum_{j=1}^{k} \alpha_{mj} = 0$ for all $m \in \{1, \ldots, \eta\}$ and $\alpha_{mj} \neq 0$ for some m and j. Then, as a result $S'\mathbf{v}_1 = -\frac{1}{k+l-1}\mathbf{v}_1$.

Therefore, there exists a set of $\eta(k-1)$ orthogonal eigenvectors $\mathbf{v}_1$, implying that $-\frac{1}{k+l-1}$ is an eigenvalue of S' with multiplicity $\eta(k-1)$.

Next, consider $\mathbf{v}_2 = [0, \ldots, 0, \alpha_{11}, \ldots, \alpha_{1k}, \alpha_{21}, \ldots, \alpha_{2k}, \ldots, \alpha_{\eta 1}, \ldots, \alpha_{\eta k}]^T$, where the first l entries are zero, be a N-dimensional vector such that for all $m \in \{1, \ldots, \eta\}$ it holds $\alpha_{mj} = \alpha_m$, $\sum_{j=1}^{\eta} \alpha_j = 0$ and $\alpha_j \neq 0$ for some j. Then, it follows that $S'\mathbf{v}_2 = \frac{k-1}{k+l-1}\mathbf{v}_2$. Therefore, there exists a set of $\eta - 1$ orthogonal eigenvectors $\mathbf{v}_2$, implying that $\frac{k-1}{k+l-1}$ is an eigenvalue of S' with multiplicity $\eta - 1$.

Now, we define the N-dimensional vector $\mathbf{v}_3 = [\beta_1, \ldots, \beta_l, 0, \ldots, 0]^T$, such that $\sum_{j=1}^{l} \beta_j = 0$ and $\beta_j \neq 0$ for some j. Hence, $S'\mathbf{v}_3 = -\frac{1}{\eta k+l-1}\mathbf{v}_3$. Therefore, there exists a set of $l-1$ orthogonal eigenvectors $\mathbf{v}_3$, implying that $-\frac{1}{\eta k+l-1}$ is an eigenvalue of S' with multiplicity $l-1$.

Next, let $\mathbf{v}_4 = [1, \ldots, 1, x, \ldots, x]^T$, where the first l entries are one, be a N-dimensional vector with $x = -\frac{l\sqrt{\eta k+l-1}}{\eta k\sqrt{k-1+l}}$. Then, it can be shown that $S'\mathbf{v}_4 = (\frac{l-1}{\eta k-1+l} - \frac{l}{k-1+l})\mathbf{v}_4$.

It follows that $\frac{l-1}{\eta k+l-1} - \frac{l}{k+l-1}$ is an eigenvalue of S'.

Finally, consider $\mathbf{v}_5 = [1, \ldots, 1, y, \ldots, y]^T$, where the first l entries are one, be a N-dimensional vector with $y = \frac{\sqrt{k+l-1}}{\sqrt{\eta k+l-1}}$. Then, it can be verified that $S'\mathbf{v}_5 = \mathbf{v}_5$.

It follows that 1 is an eigenvalue of S'.

Because the sum of the multiplicities of the found eigenvalues, i.e. $\eta(k-1)+(\eta-1)+l-1+1+1$ equals the number of nodes $\eta k + l$, we have found all eigenvalues of S'. This finishes the proof.

Appendix B. Proof of Lemma 7

Let $\mathbf{w}_1 = [0, \ldots, 0, \alpha_{11}, \ldots, \alpha_{1k}, \alpha_{21}, \ldots, \alpha_{2k}, \ldots, \alpha_{\eta 1}, \ldots, \alpha_{\eta k}]^T$, where the first l entries are zero, be a N-dimensional vector such that $\sum_{j=1}^{k} \alpha_{mj} = 0$ for all $m \in \{1, \ldots, \eta\}$ and $\alpha_{mj} \neq 0$ for some m and j. Then, as a result $S''\mathbf{w}_1 = -\frac{1}{k+l-1}\mathbf{w}_1$. Therefore, there exists a set of $\eta(k-1)$ orthogonal eigenvectors $\mathbf{w}_1$, implying that $-\frac{1}{k+l-1}$ is an eigenvalue of S'' with multiplicity $\eta(k-1)$.

Next, consider $\mathbf{w}_2 = [0, \ldots, 0, \alpha_{11}, \ldots, \alpha_{1k}, \alpha_{21}, \ldots, \alpha_{2k}, \ldots, \alpha_{\eta 1}, \ldots, \alpha_{\eta k}]^T$, where the first l entries are zero, be a N-dimensional vector such that for all $m \in \{1, \ldots, \eta\}$ it holds $\alpha_{mj} = \alpha_m$, $\sum_{j=1}^{\eta} \alpha_j = 0$ and $\alpha_j \neq 0$ for some j. Then, it follows that $S''\mathbf{w}_2 = \frac{k-1}{k+l-1}\mathbf{w}_2$.

Therefore, there exists a set of $\eta - 1$ orthogonal eigenvectors $\mathbf{w}_2$, implying that $\frac{k-1}{k+l-1}$ is an eigenvalue of S'' with multiplicity $\eta - 1$.

Now, we define the N-dimensional vector $\mathbf{w}_3 = [\beta_1, \ldots, \beta_l, 0, \ldots, 0]^T$, such that $\sum_{j=1}^{l} \beta_j = 0$ and $\beta_j \neq 0$ for some j. Hence,

$$S''\mathbf{w}_3 = 0\mathbf{w}_3.$$

Therefore, there exists a set of $l-1$ orthogonal eigenvectors $\mathbf{w}_3$, implying that 0 is an eigenvalue of S'' with multiplicity $l-1$.

Next, let $\mathbf{w}_4 = [1, \ldots, 1, x', \ldots, x']^T$, where the first l entries are one, be a N-dimensional vector with $x' = -\frac{l}{\sqrt{\eta k}\sqrt{k-1+l}}$. Then, it can be seen that $S''\mathbf{w}_4 = -\frac{l}{k-1+l}\mathbf{w}_4$.

It follows that $-\frac{l}{k+l-1}$ is an eigenvalue of S''.

Finally, consider $\mathbf{w}_5 = [1, \ldots, 1, y', \ldots, y']^T$, where the first l entries are one, be a N-dimensional vector with $y' = \frac{\sqrt{k+l-1}}{\sqrt{\eta k}}$. Then, it can be verified that $S''\mathbf{w}_5 = \mathbf{w}_5$.

It follows that 1 is an eigenvalue of S''.

Because the sum of the multiplicities of the found eigenvalues, i.e. $\eta(k-1)+(\eta-1)+l-1+1+1$ equals the number of nodes $\eta k + l$, we have found all eigenvalues of S''. This finishes the proof.

Note that the symmetric matrices S' and S'' both have eigenvalues $-\frac{1}{k+l-1}$ and $\frac{k-1}{k+l-1}$, with the same corresponding eigenvectors.

References

[1] F.K. Bell, A note on the irregularity of graphs, Linear Algebra Appl. 161 (1992) 45–54.
[2] E.R. van Dam, R.E. Kooij, The minimal spectral radius of graphs with a given diameter, Linear Algebra Appl. 423 (2) (2007) 408–419.
[3] E. Estrada, Quantifying network heterogeneity, Phys. Rev. E 82 (2010) 066102.
[4] E. Estrada, When local and global clustering of networks diverge, Linear Algebra Appl. 488 (2016) 249–263.
[5] E. Estrada, M. Benzi, Core–satellite graphs: Clustering, assortativity and spectral properties, Linear Algebra Appl. (ISSN: 0024-3795) 517 (2017) 30–52.
[6] J.J. Hunter, The role of Kemeny's constant in properties of Markov chains, Comm. Statist. Theory Methods 43 (7) (2014) 1309–1321.
[7] J.G. Kemeny, J.L. Snell, Finite Markov Chains, D. Van Nostrand, Princeton, NJ, 1960.
[8] S. Kirkland, Z. Zeng, Kemeny's constant and an analogue of Braess' paradox for trees, Electron. J. Linear Algebra 31 (1) (2016) 444–464.
[9] S. Knight, H.X. Nguyen, N. Falkner, R. Bowden, M. Roughan, The internet topology zoo, IEEE J. Sel. Areas Commun. 29 (9) (2011) 1765–1775.
[10] R.E. Kooij, On generalized windmill graphs, Linear Algebra Appl. 565 (2019) 25–46.
[11] M. Levene, G. Loizou, Kemeny's constant and the random surfer, Amer. Math. Monthly 109 (8) (2002) 741–745.
[12] L. Lovász, Random walks on graphs: A survey, in: Paul Erdös is Eighty, vol. 2, Bolyai Society, Mathematical Studies, Keszthely, Hungary, 1993, pp. 1–46.
[13] J.L. Palacios, On the Kirchhoff index of regular graphs, Int. J. Quantum Chem. 110 (7) (2010) 1307–1309.
[14] J.L. Palacios, J.M. Renom, Bounds for the Kirchhoff index of regular graphs via the spectra of their random walks, Int. J. Quantum Chem. 110 (9) (2010) 1637–1641.
[15] J.L. Palacios, J.M. Renom, Broder and Karlin's formula for hitting times and the Kirchhoff index, Int. J. Quantum Chem. 111 (1) (2011) 35–39.
[16] R. Patel, P. Agharkar, F. Bullo, Robotic surveillance and Markov chains with minimal weighted Kemeny constant, IEEE Trans. Automat. Control 60 (2015) 3156–3167.
[17] P. Van Mieghem, Graph Spectra for Complex Networks, Cambridge University Press, 2011.
[18] L. Von Collatz, U. Sinogowitz, Spektren endlicher grafen, Abh. Math. Semin. Univ. Hambg. 21 (1) (1957) 63–77.
[19] X. Wang, J.L.A. Dubbeldam, P. Van Mieghem, Kemeny's constant and the effective graph resistance, Linear Algebra Appl. 535 (2017) 231–244.