APPROXIMATIONS FOR KEMENY'S CONSTANT FOR SEVERAL FAMILIES OF GRAPHS AND REAL-WORLD NETWORKS

Robert Kooij

5 December 2024

Graphs&Data@TUDelft

Joint work with ***Johan Dubbeldam (DIAM)***

TU Delft

Introduction

- Faculty: EEMCS
- Department: Quantum & Computer Engineering
- Section: **Network Architectures & Services**

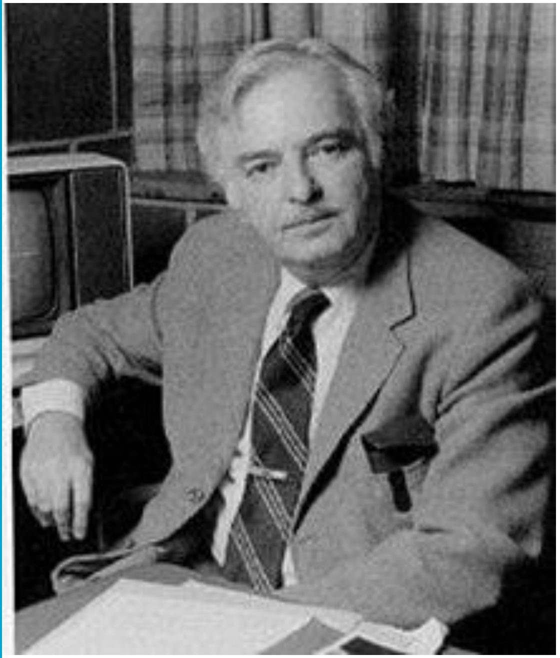
- Quantify robustness of a network
  - use network metrics

S. Freitas, D. Yang, S. Kumar, H. Tong and D. H. Chau, "Graph Vulnerability and Robustness: A Survey," in *IEEE Transactions on Knowledge and Data Engineering*, doi: 10.1109/TKDE.**2022**.3163672.

- 151 references, including 7 papers from NAS

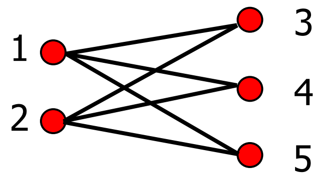
Aim of this talk: discuss **Kemeny's Constant**

John George Kemeny



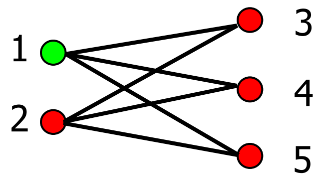
- János György Kemény (1926 - 1992)
- Joined Manhattan Project aged 18!
- Invented BASIC programming language

What is Kemeny's Constant?



- Undirected graphs
- Random walks

What is Kemeny's Constant?



- Undirected graphs
- Random walks

Pr(1→3) = 1/3
Pr(1→4) = 1/3
Pr(1→5) = 1/3

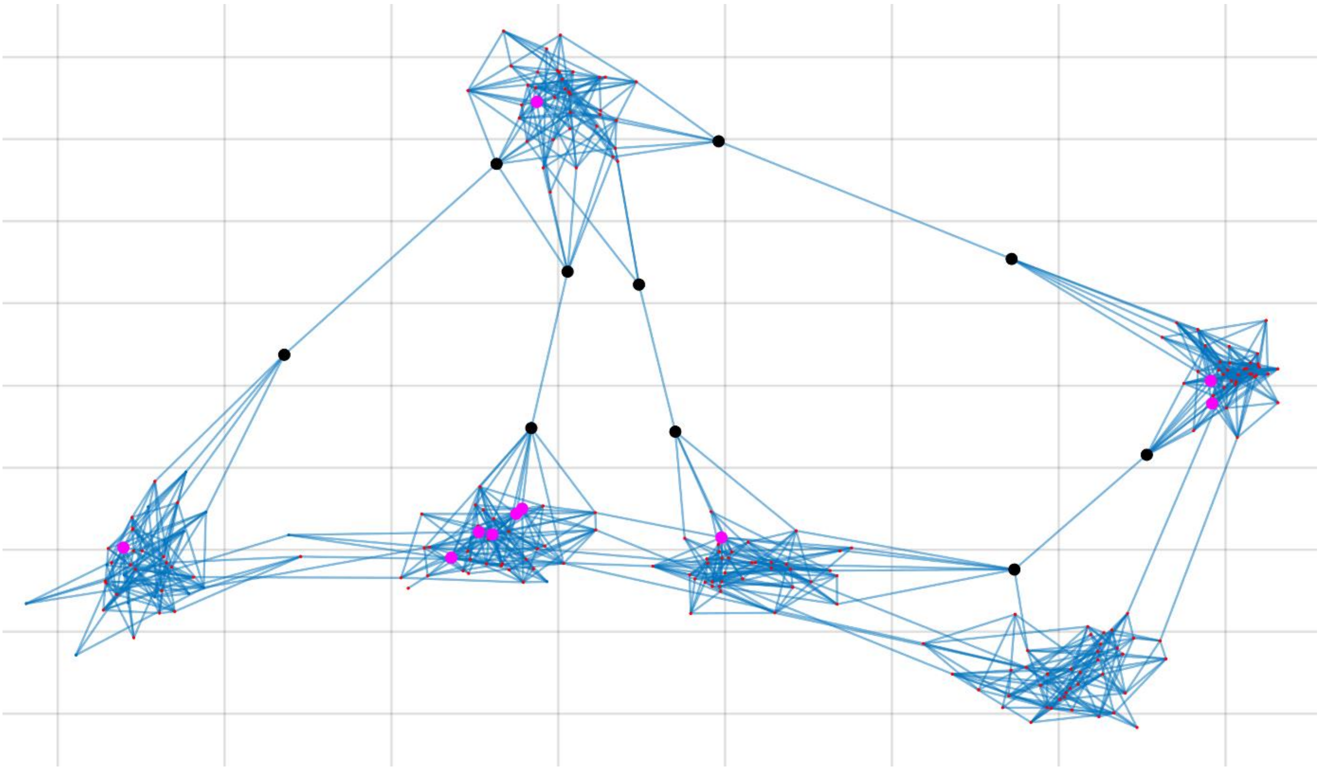
- Mean first passage time

$$m_{13} = 1 \cdot \frac{1}{3} + 3 \cdot \frac{2}{3} \cdot \frac{1}{3} + 5 \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} + \ \dots \ = 5$$

- Kemeny's Constant = mean first passage time from a given node to a "randomly" chosen other node

Applications

- Efficient testing for COVID in communities



PLOS ONE: 2020

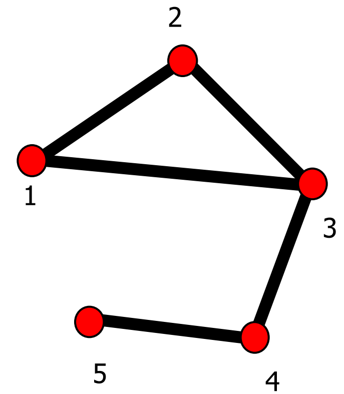
Background

- Joint work with Johan Dubbeldam (EEMCS/DIAM)

Discrete Applied Mathematics 285 (2020) 96–107

- Maria Predari

Background



- A: adjacency matrix of graph G(N,L)
- Δ: diagonal degree matrix

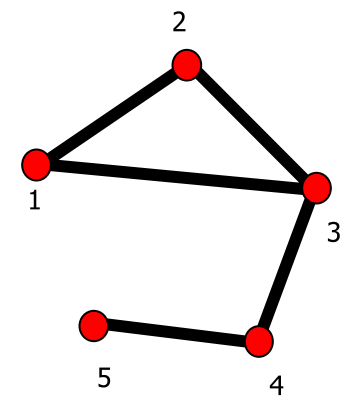
$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \qquad \Delta = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- P: transition probability matrix $P = \Delta^{-1} A$

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \qquad \pi = \begin{pmatrix} 1/5 \\ 1/5 \\ 3/10 \\ 1/5 \\ 1/10 \end{pmatrix}$$

- π: steady state probability vector $\pi^T P = \pi^T$

Background



- Q: Laplacian matrix $Q = \Delta - A$

$$Q = \begin{pmatrix} 2 & -1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ -1 & -1 & 3 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

- eigenvalues: $o = \mu_N \leq \mu_{N-1} \leq \cdots \leq \mu_1$

- $Q^\dagger$: Moore-Penrose pseudo inverse of Laplacian

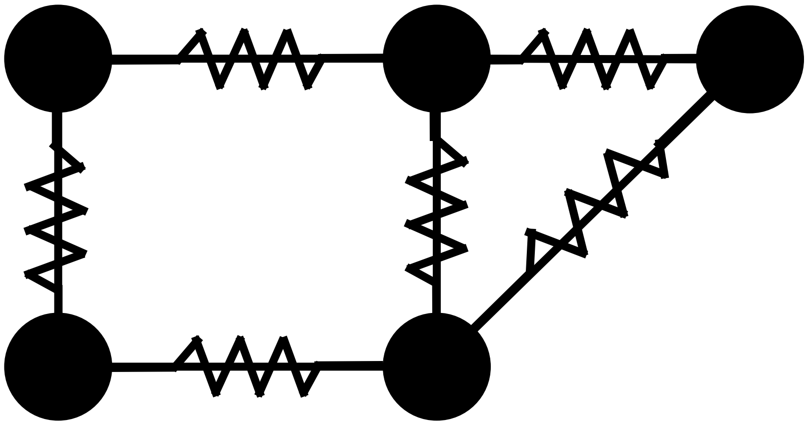
$$Q^\dagger = (Q + \frac{J}{N})^{-1} - \frac{J}{N}$$

TUDelft

Background

- R_G: effective graph resistance

Effective resistance between two nodes



Considers all paths between the nodes

R_G: Sum of all effective resistances

$$R_G = N \sum_{i=1}^{N-1} \frac{1}{\mu_i}$$

$$R_G = N trace(Q^{\dagger})$$

Background

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

m_{ji}: mean first passage time

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

λ_k: eigenvalues of $\Delta^{-1/2} A \Delta^{-1/2}$

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

ζ: vector with diagonal elements $Q^\dagger$

d: degree vector

Wang, Dubbeldam, Van Mieghem (LAA;2017)

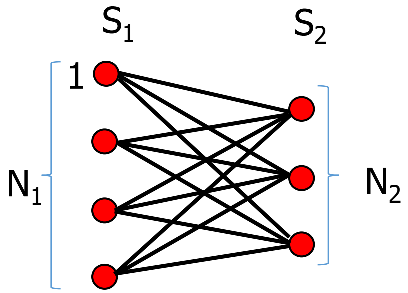
TUDelft

Overview

- Complete bi-partite graphs and some trees
- Windmill graphs
- Relation with effective graph resistance
- A sharp upper bound
- Real-world networks
- A class of biregular graphs
- Analysis for large networks
- Future work
- Wrap-up

TU Delft

Complete bi-partite graphs and some trees

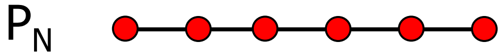


K_{N_1,N_2}

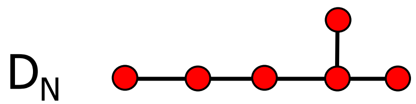
Using random walks and probabilities

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

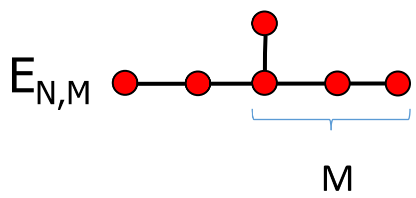
Complete bi-partite graphs and some trees



$$K(p) = \frac{1}{3}N^2 - \frac{2}{3}N + \frac{1}{2}$$



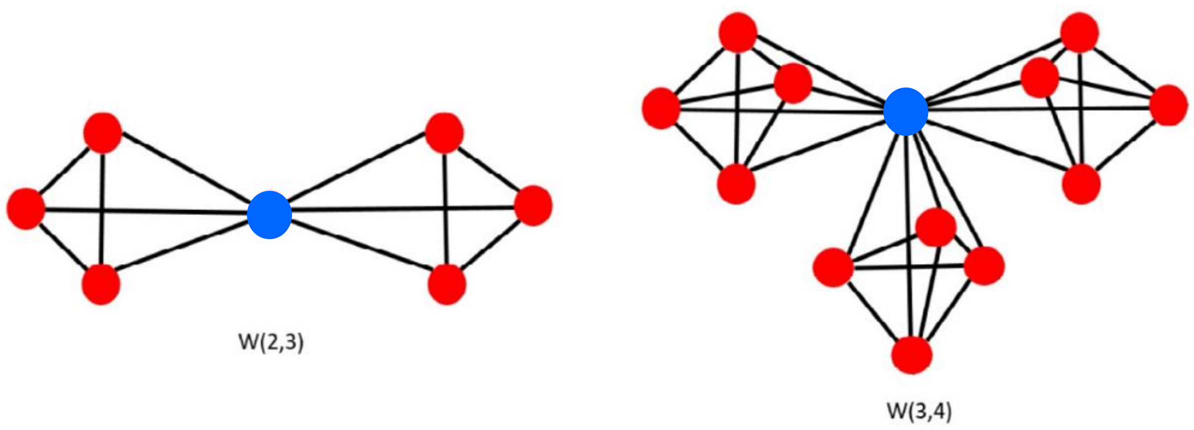
$$K(p) = \frac{1}{3}N^2 - \frac{2}{3}N + \frac{1}{2} - \frac{2(N-3)}{N-1}$$



$$K(p) = \frac{1}{3}N^2 - \frac{2}{3}N + \frac{1}{2} - \frac{2(M(N-3) - M(M-1))}{N-1}$$

Windmill graphs

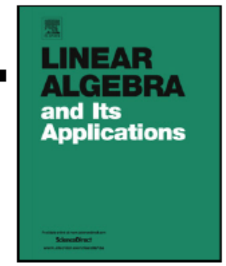
- Introduced by Estrada (2016)
  - η cliques of order k
  - connected through central node



- N → ∞
  - clustering coefficient → 1 while transitivity index → 0

Windmill graphs

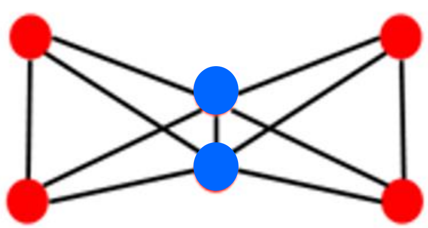
Linear Algebra and its Applications 565 (2019) 25–46



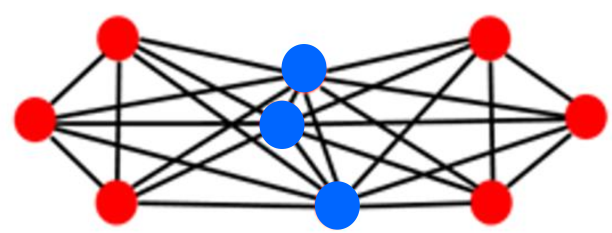
On generalized windmill graphs

Robert Kooij [a,b,*]

- Generalized windmill graph: W'(η,k,l)
  - η cliques of order k
  - connected through central clique of order l



W'(2,2,2)

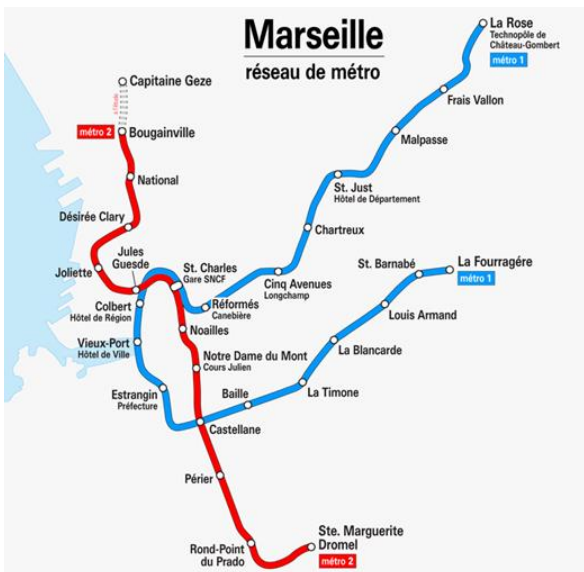


W'(2,3,3)

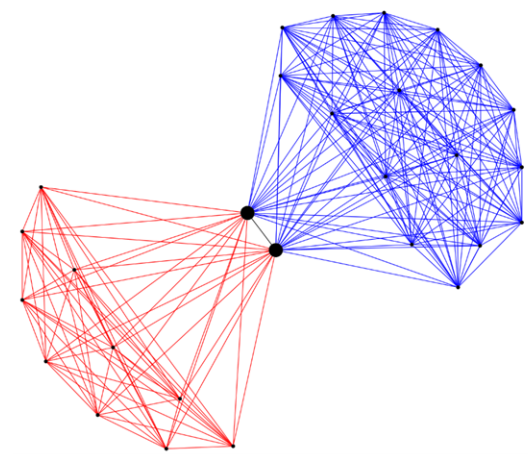
TUDelft

Windmill graphs

- Application: Public Transport Networks



- Representation in P-space



- Prof. Oded Cats
- Fac. CiTG

Windmill graphs

- K(P) for W'(η,k,l) using spectrum of $\Delta^{-1/2} A \Delta^{-1/2}$

spectrum: $$\{(-\frac{1}{k+l-1})^{\eta(k-1)}, (\frac{k-1}{k+l-1})^{\eta-1}, (-\frac{1}{\eta k+l-1})^{l-1}, (\frac{l-1}{\eta k+l-1} - \frac{l}{k+l-1})^{1}, (1)^{1}\}$$

Windmill graphs

W′(η,k,l)

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

Special case: l = 1; windmill graph W(η,k)

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

Relation with effective graph resistance

- Palacios (2010)
  - r-regular graphs on N nodes

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

- Approximation for non-regular graphs
  - Average degree D
  - Heterogeneity index H

$$D = \frac{2L}{N}$$

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

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

- Choose f(N,L) such that K*(P) is exact for K_{N_1,N_2}

TUDelft

Relation with effective graph resistance

- for $K_{N1,N2}$

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

$$D = \frac{2N_1N_2}{N_1 + N_2} \qquad H = \frac{N_1N_2(N_1 - N_2)^2}{(N_1 + N_2)^2} \qquad R_G = (N_1 + N_2)(\frac{N_2 - 1}{N_1} + \frac{N_1 - 1}{N_2} + \frac{1}{N_1 + N_2})$$

$$f = \frac{1 - 2N_1 - 2N_2}{2N_1N_2}$$

- for $K_{N1,N2}$

$$N = N_1 + N_2$$

$$L = N_1N_2$$

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

Relation with effective graph resistance

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

- Exact for regular graphs
- Exact for complete bipartite graphs
- Also exact for windmill graphs!

TUDelft

Relation with effective graph resistance

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

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

TU Delft

A sharp upper bound

- K*(P) is not an upper bound

$$K(P) = \zeta^T d - \frac{d^T Q^\dagger d}{2L} \leq \zeta^T d - \frac{H}{D\mu_1} \equiv K_U(P)$$

Wang, Dubbeldam, Van Mieghem (LAA;2017)

TU Delft

A sharp upper bound

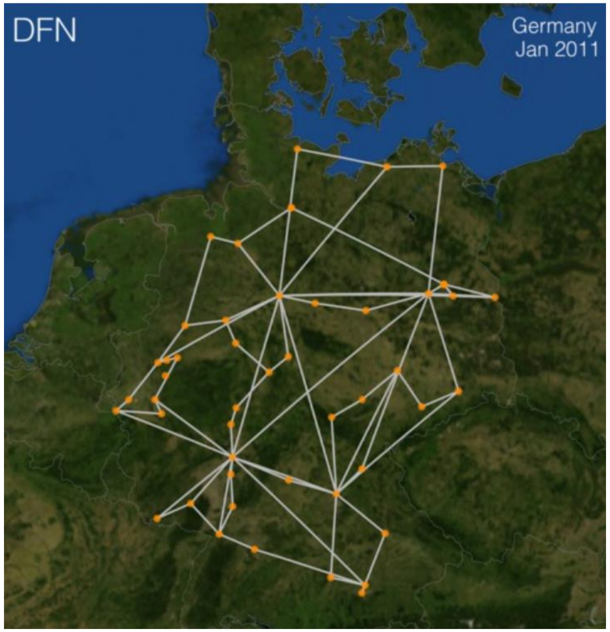
Table 2
Kemeny's constant and the upper bound $K_U(P)$, for several graphs.

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

- $K_U(P)$ is tight for (generalized) windmill graphs

Real-world networks

- Data from Internet Topology Zoo
  - 243 communication networks



TUDelft

Real-world networks

Table 3
Kemeny's constant, the approximations 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_U
Arpanet196912	4	4	2.54	2.73	2.60
Renam	5	4	3.50	3.50	3.50
Mren	6	5	4.50	4.50	4.50
UsCarrier	158	189	1175.99	1265.48	1176.68
Cogentco	197	245	1082.45	1197.24	1083.35
Kdl	754	899	5907.29	6264.78	5908.32

TU Delft

Real-world networks

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

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

TUDelft

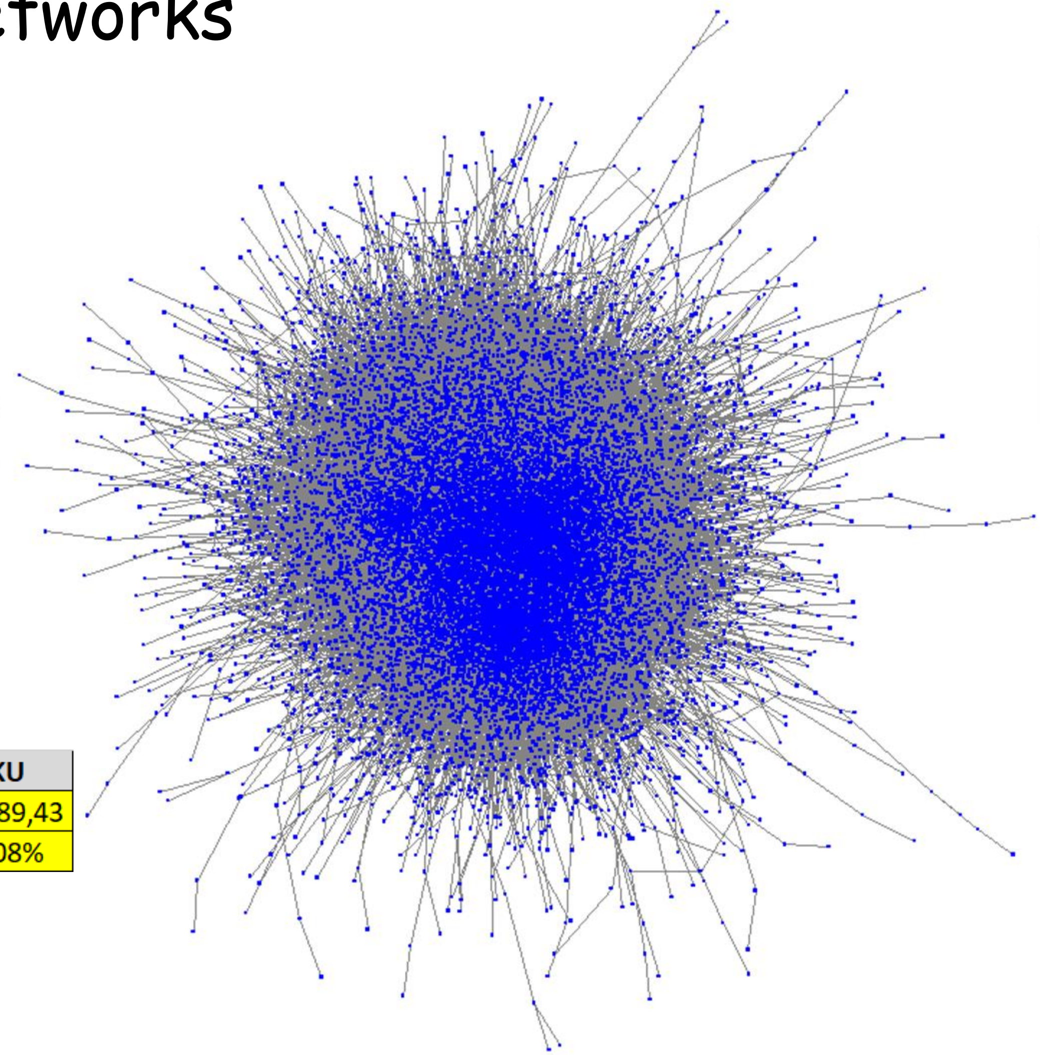
Real-world networks

bio-CE-CX.edges (undirected)

Summary Statistics

Number of nodes	15063
Number of edges	245862

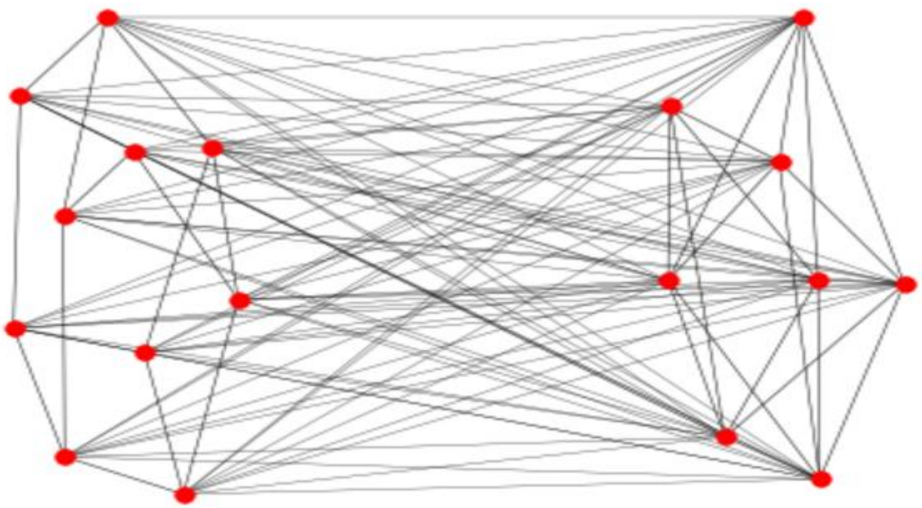
Network	K	K*	KU
bio-CE-CX	17375,46	128159,7	17389,43
	Rel. error	638%	0,08%



TUDelft

A class of biregular graphs

$k_1 = 3$
$N_1 = 10$



$k_2 = 5$
$N_2 = 8$

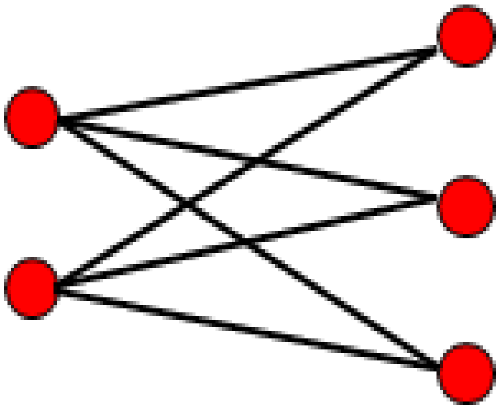
G_1
k_1-regular graph
N_1 nodes

join

G_2
k_2-regular graph
N_2 nodes

TU Delft

A class of biregular graphs



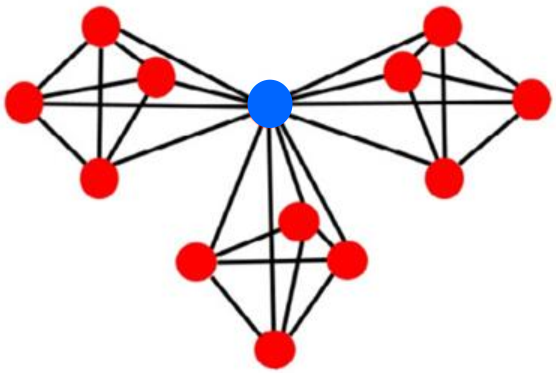
complete bi-partite graph

G_1
$k_1 = 0$
N_1 nodes

join

G_2
$k_2 = 0$
N_2 nodes

A class of biregular graphs



windmill graph

G_1
η copies of K_m

join

G_2
1 node

A class of biregular graphs

Theorem: $K_U(P)$ is tight for this class of biregular graphs

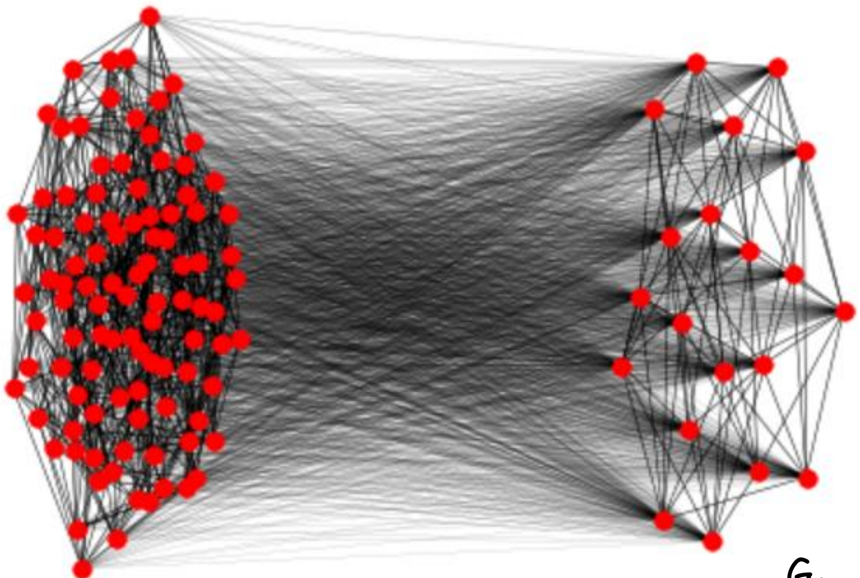
Proof: exploit properties of eigenvalues and eigenvectors of Q

Further exploration of an upper bound for Kemeny's Constant

Robert E. Kooij[a,b], Johan L.A. Dubbeldam[a], Maria Predari[c]

TU Delft

A class of biregular graphs



G_1
$k_1 = 10$
$N_1 = 100$

$K = K_U = 119.24$

G_2
$k_2 = 8$
$N_2 = 20$

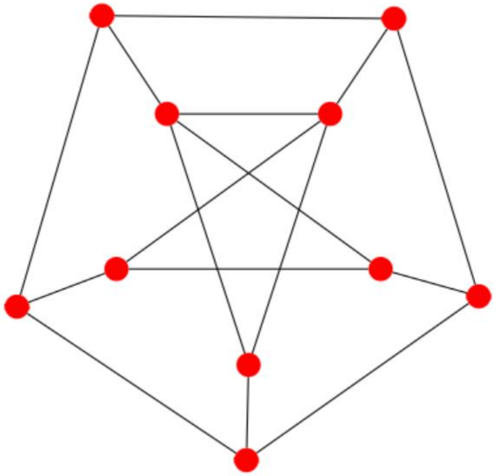
TU Delft

A class of biregular graphs

- Properties of our class of graphs
  - Biregular
  - Diameter = 2
- Not sufficient for K_U to be tight

$K = 9.766$

$K_U = 9.774$



Analysis for large networks

$$K(P) = \zeta^T d - \frac{d^T Q^\dagger d}{2L} \leq \zeta^T d - \frac{H}{D\mu_1} \equiv K_U(P)$$

$O(N^3)$

Only diagonal elements of $Q^\dagger$

E. Angriman et al. Approximation of the diagonal of a Laplacian's pseudoinverse for complex network analysis, in: ESA, Vol. 173 of LIPIcs, 2020, pp. 6:1–6:24.

- Fast approximation of diagonal elements of $Q^\dagger$
- Errors due to sampling parameter ε

Analysis for large networks

Graph	Type	ID	$\|V\|$	$\|E\|$
inf-power	infrastructure	inf	4K	6K
facebook-ego-combined	social	fac	4K	8.8K
p2p-Gnutella04	internet	p2p	10K	39K
ca-HepPh	collaboration	ca-	11K	117K
arxiv-astro-ph	collaboration	arx	17K	196K
eat	words	eat	23K	297K
arenas-pgp	infrastructure	are	24K	10K
as-caida20071105	internet	as-	26K	53K
ia-email-EU	communication	ia-	32K	54.4K
loc-brightkite	social	lob	57K	213K
soc-Slashdot0902	social	soc	82K	504K
flickr	images	fli	106K	2.31M
livemocha	social	liv	104K	2.19M
loc-gowalla-edges	social	log	196K	950K
web-NotreDame	web	web	325K	1.09M
citeseer	citation	cit	365K	1.72M

medium

large

Analysis for large networks

Results for medium graphs

ε	Average Rel. Error	Average Speed Up
0.1	0.33%	2
0.3	0.27%	18
0.5	0.25%	48
0.9	1.26%	141

TUDelft

Analysis for large networks

Results for large graphs

Graph	$K_U(P)$	Time (in sec)
lob	80,903	48.83
soc	96,102	50.87
fli	122,185	98.11
liv	120,525	37.07
log	271,577	310.77
web	1,009,760	4,279.36
cit	508,244	4,571.51

TU Delft

Future work

- Use K as proxy for network robustness
- Use K as indicator for detecting anomalies in bank transaction networks
- Define quantum Kemeny's constant

Wrap-up

- Explicit expressions for K for some graph families
- Approximation using effective graph resistance
- Sharp upper bound
- Validation on real-world networks
- New class of graphs for which upper bound is sharp
- Analysis of some large networks

Thanks for your attention!

- r.e.kooij@tudelft.nl