Random walks centrality measures and community detection

Jeremi K. Ochab, Z. Burda
Marian Smoluchowski Institute of Physics, Jagiellonian University

CompPhys12

29 Nov. 2012, Leipzig

Centrality measures

- Web search → PageRank, HITS
- Social → importance/popularity
- Transport → facility location

Community detection

- Information → topic related www, routing
- Social → group stability, collaboration
- Biological → metabolic, protein networks

Centrality measures

- Web search → … random surfer?
- Social → … communication, phone calls
- Transport → … cars

WHAT PROCESS?

Community detection

- Information → … routing tables
- Social → …
- Biological → … activated/repressed expression

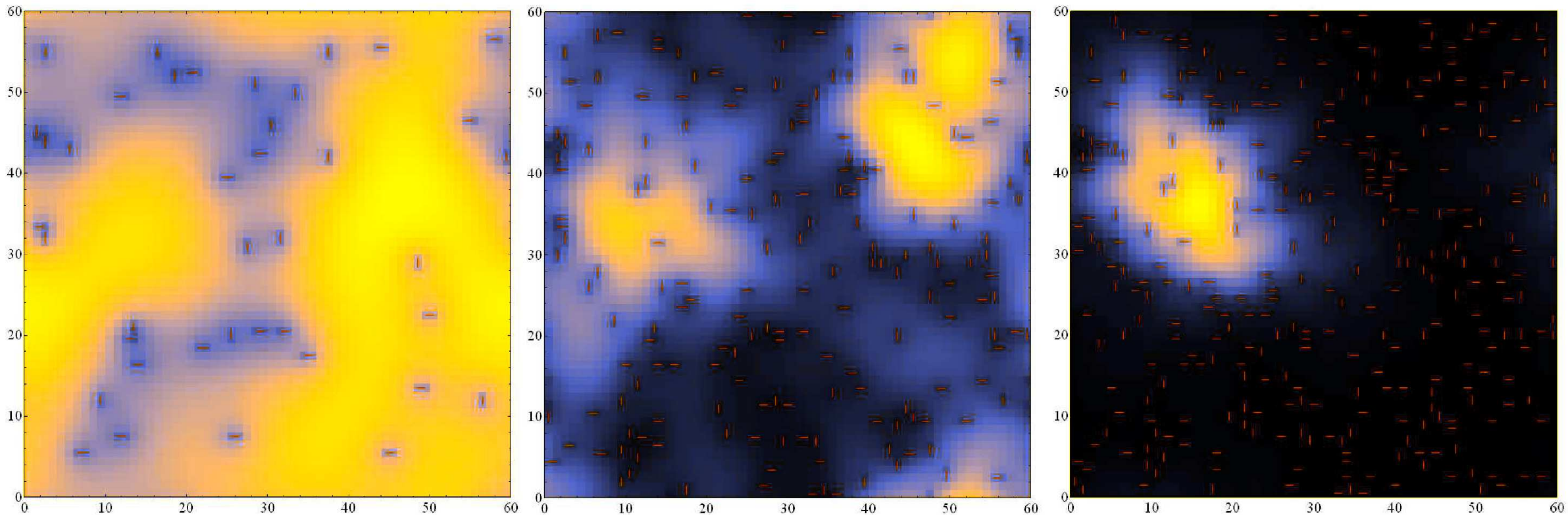
Random Walks compared

	GENERIC RW	MAXIMAL-ENTROPY RW
TRANSITION MATRIX	$P_{ba} = \frac{A_{ba}}{k_a}$	$P_{ba} = \frac{A_{ba}}{\lambda_0}\frac{\psi_{0b}}{\psi_{0a}}$
STATIONARY STATE	$\pi_a = \frac{k_a}{2L}$	$\pi_a = {\psi_{0a}}^2$
PATH PROBABILITY	$P(\gamma_{a_t a_0}) = \prod_{i=0}^{t-1}\frac{1}{k_i}$	$P(\gamma_{a_t a_0}) = \frac{1}{{\lambda_0}^t}\frac{\psi_{0a_t}}{\psi_{0a_0}}$

$\mathbf{A}$	adjacency matrix of the graph, $A_{ba} = 0$ or 1.
$\lambda_i, \vec{\psi}_i$	eigenvalues and eigenvectors of $\mathbf{A}$, $\sum_b A_{ba}\psi_{ib} = \lambda_i\psi_{ia}, \sum_b {\psi_{ib}}^2 = 1$.
P_{ba}	transition probability of a walker taking a step from node a to node b.

Z. Burda, J. Duda, J.M. Luck, and B. Waclaw, Phys. Rev. Lett. **102** (2009) 160602
Z. Burda, J. Duda, J.M. Luck, and B. Waclaw, Acta Phys. Pol. B **41** (2010) 949

MERW: localisation

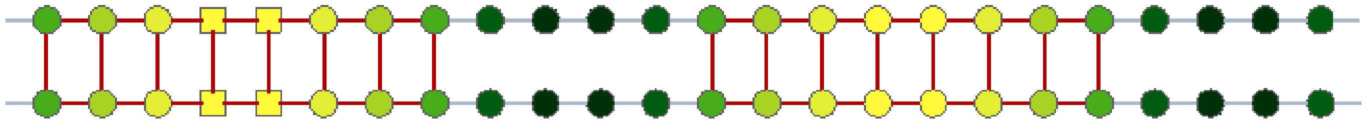


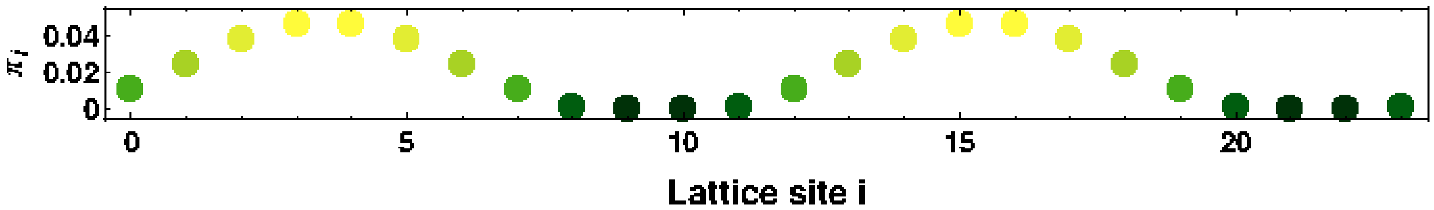
Localisation of the **stationary probability** distribution
in the largest spherical Lifshitz region.

The Wolfram Demonstrations Project http://demonstrations.wolfram.com:

- "Generic Random Walk and Maximal Entropy Random Walk"

MERW: dynamics





The relaxation time for MERW is exponential in the gap size g:

$$\tau = \exp\left(\left(f\ N^{-1/\nu} - c_\infty\right) \cdot g\right)$$

The relaxation time for GRW is independent of the gap size:

$$\tau \sim N^d$$

J.K. Ochab, Acta Phys. Pol. B 43 (2012) 1143

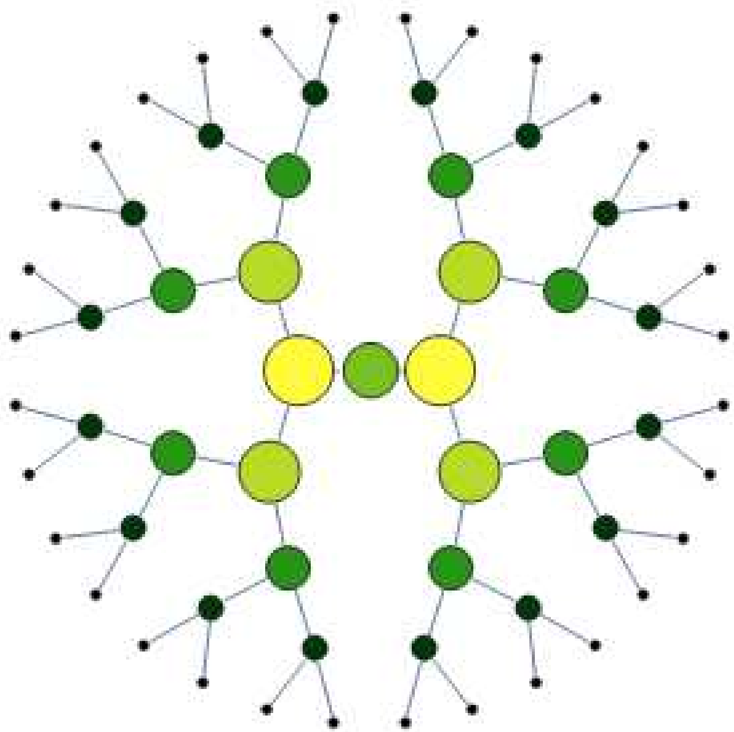
MERW: dynamics

Fast relaxation for MERW:

$$\tau \sim \ln N$$

Slow relaxation for GRW:

$$\tau \sim N$$



J.K. Ochab, Z. Burda, Phys. Rev. E 85, 021145 (2012)

The Wolfram Demonstrations Project http://demonstrations.wolfram.com :
- "Dynamics of Maximal Entropy Random Walk and Generic Random Walk on Cayley Trees"
- "Stationary States of Maximal Entropy Random Walk and Generic Random Walk on Cayley Trees"

MERW: mean first-passage times

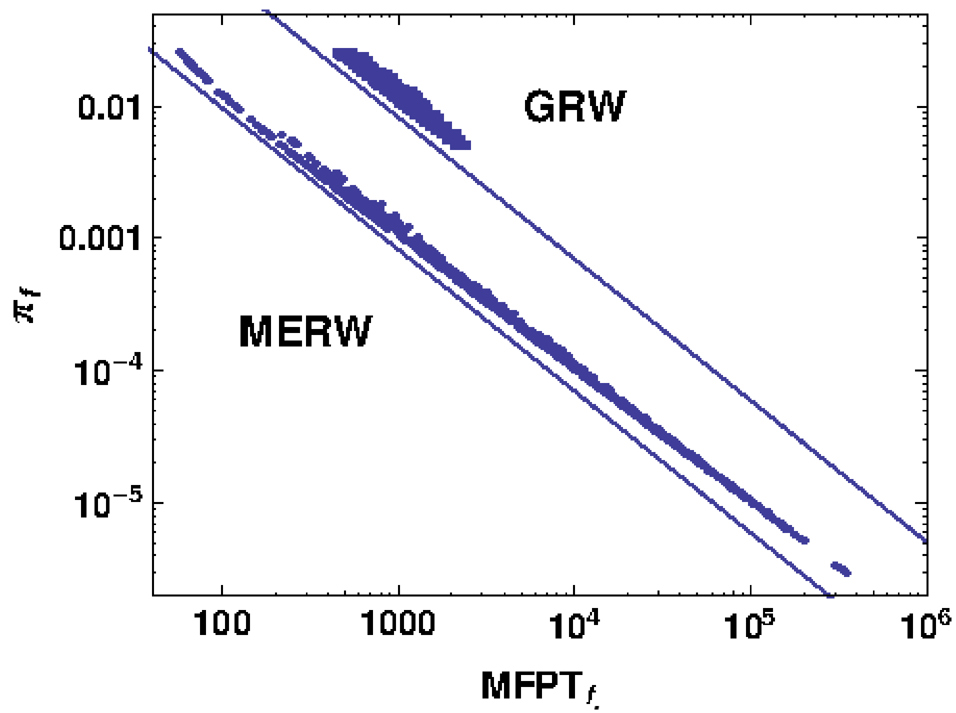
$Z = (1 - P + e\pi T)^{-1}$ = fundamental matrix ≈

$1 + P + P^2 + \cdots$

$M = (EZ_{dg} - Z)D$ = MFPT matrix

- $e = (1,1,1,\dots)^T$
- $(Z_{dg})_{ii} = Zii$
- $(D)_{ii} = 1/\pi_i$

$M_{if} \sim 1/\pi_f$

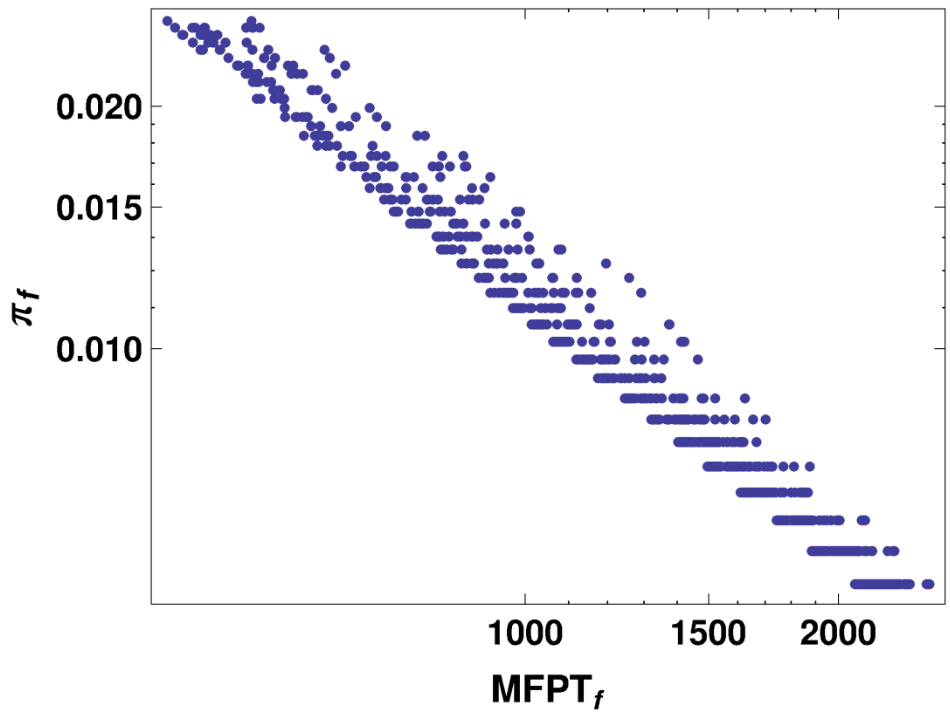


MERW: mean first-passage times

$Z = (1 - P + e\pi T)^{-1}$ = fundamental matrix ≈

$1 + P + P^2 + \cdots$

$M = (EZ_{dg} - Z)D$ = MFPT matrix



- $e = (1,1,1,\ldots)^T$
- $(Z_{dg})_{ii} = Zii$
- $(D)_{ii} = 1/\pi_i$

$M_{if} \sim 1/\pi_f$

J.K. Ochab, arXiv:1206.4094 [physics.soc-ph]

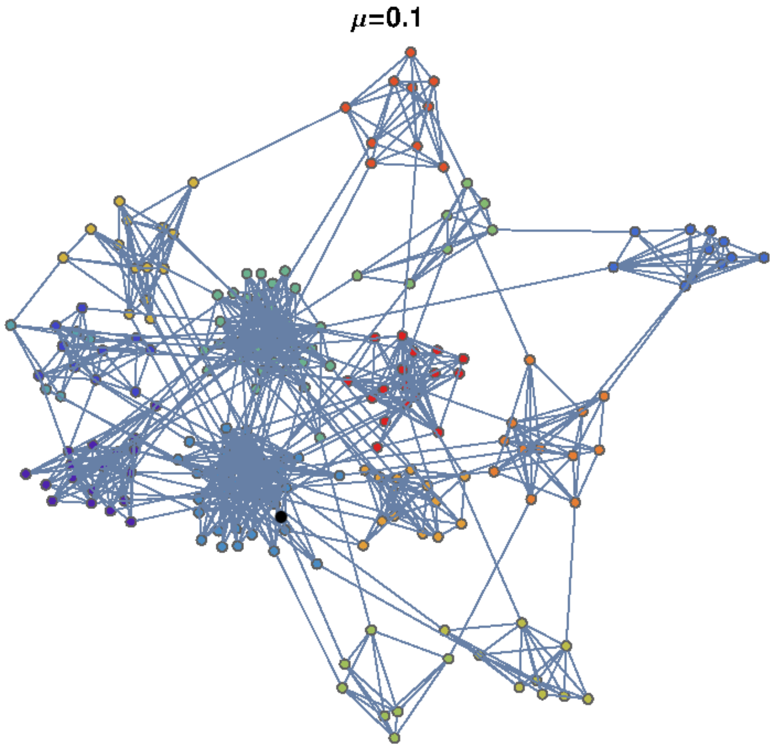
CENTRALITY MEASURES

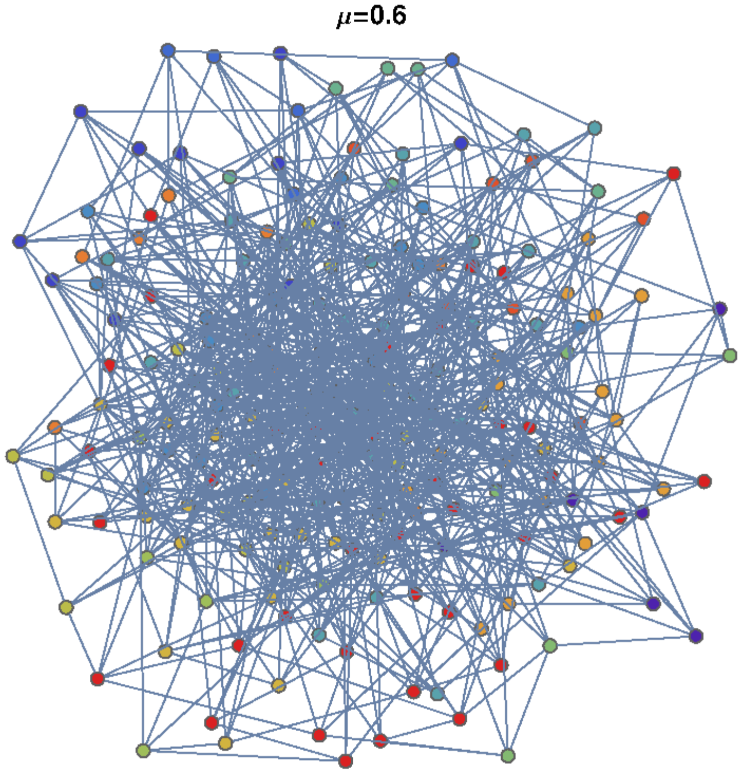
Benchmark graphs

Power-law **degree** dist., <k>=const.

Power-law **community size** dist., predefined community assignement

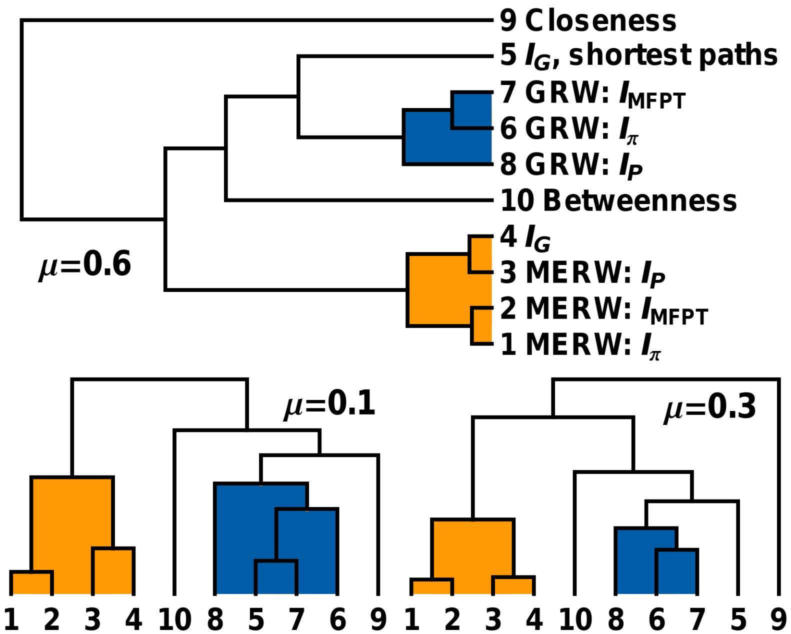
μ – mixing parameter = fraction of a node's links outside its group





A. Lancichinetti, S. Fortunato, F. Radicchi, Phys. Rev. E 78 (2008) 046110

Comparison



EIGENVECTOR	$\vec{\mathbf{I}}_\psi = \vec{\psi}_0$
STATIONARY STATE	$\vec{\mathbf{I}}_\pi = \vec{\pi}$
WEIGHTED PATHS	$\vec{\mathbf{I}}_G = (\mathbf{G}(\lambda) - \mathbf{1})\vec{\mathbf{e}}$
STOCHASTIC MATRIX	$\vec{\mathbf{I}}_P = \sum_{t=1}^{T} \vec{\pi}(0)^T \mathbf{P}^t$
MFPT	$\vec{\mathbf{I}}_{MFPT} = N / \sum_i M_{if}$

The dendrograms show affinity between different centrality measures.

- **Blue** group corresponds to GRW
- **Orange** group corresponds to measures unified by MERW

Orange members are on average closer to each other and further away from other measures than **Blue**.

J.K. Ochab, arXiv:1206.4094 [physics.soc-ph]

J.K. Ochab, Z. Burda, arXiv:1208.3688 [physics.soc-ph]

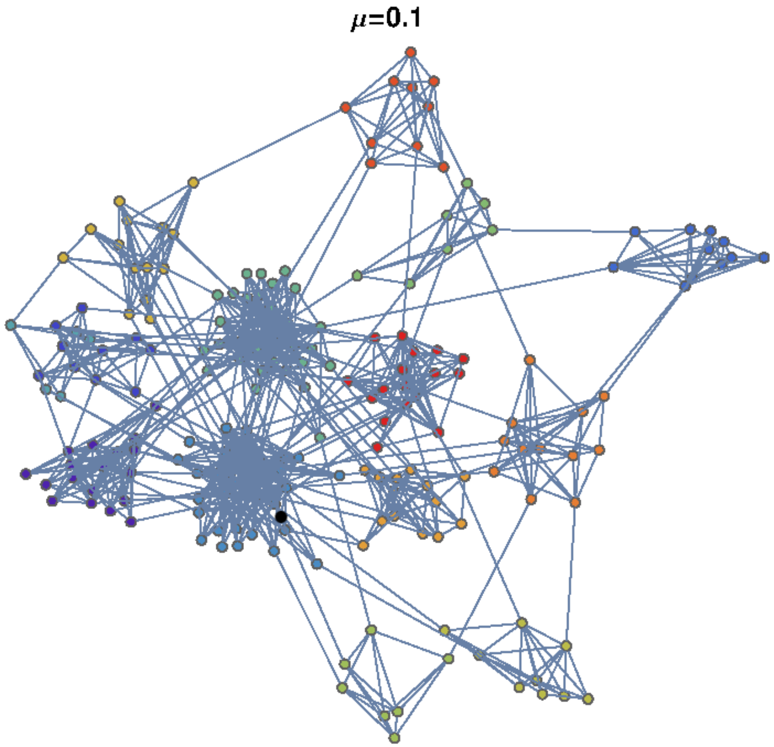
COMMUNITY DETECTION

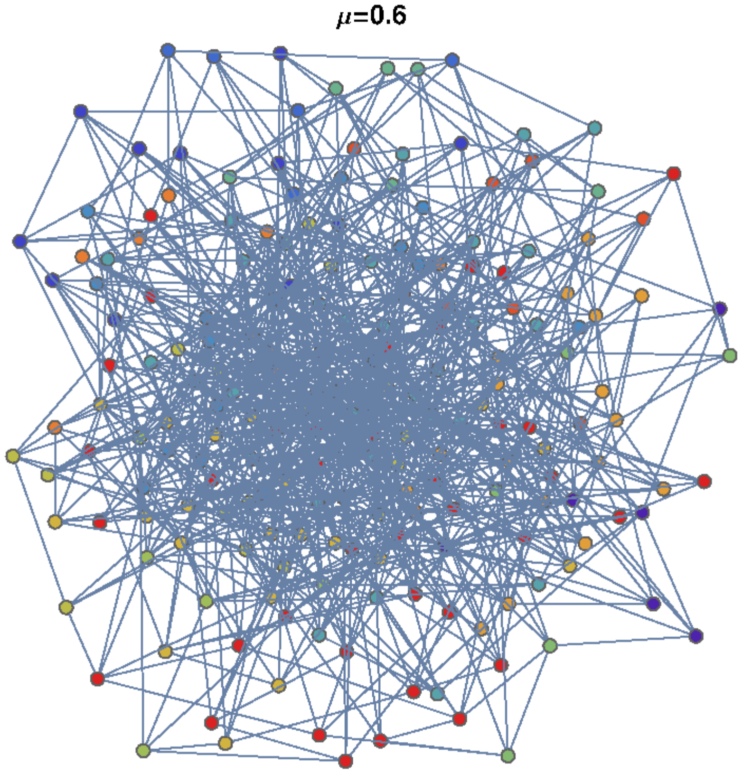
Benchmark graphs

Power-law **degree** dist., <k>=const.

Power-law **community size** dist., predefined community assignement

μ – mixing parameter = fraction of a node's links outside its group





A. Lancichinetti, S. Fortunato, F. Radicchi, Phys. Rev. E 78 (2008) 046110

Normalised Mutual Information

Measures independence of distributions → dissimilarity of graph partitions

$$NMI(X;Y) = \frac{I(X;Y)}{\sqrt{H(X)H(Y)}}$$

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p_1(x)p_2(y)}\right)$$

$$H(X) = \sum_{x \in X} p(x) \log p(x)$$

$$H(\pi^a) = \sum_{h}^{k^{(a)}} \frac{n_h^a}{n} \log(\frac{n_h^a}{n})$$

$$H(\pi^b) = \sum_{\ell}^{k^{(b)}} \frac{n_\ell^b}{n} \log(\frac{n_\ell^b}{n})$$

π – partition into communities

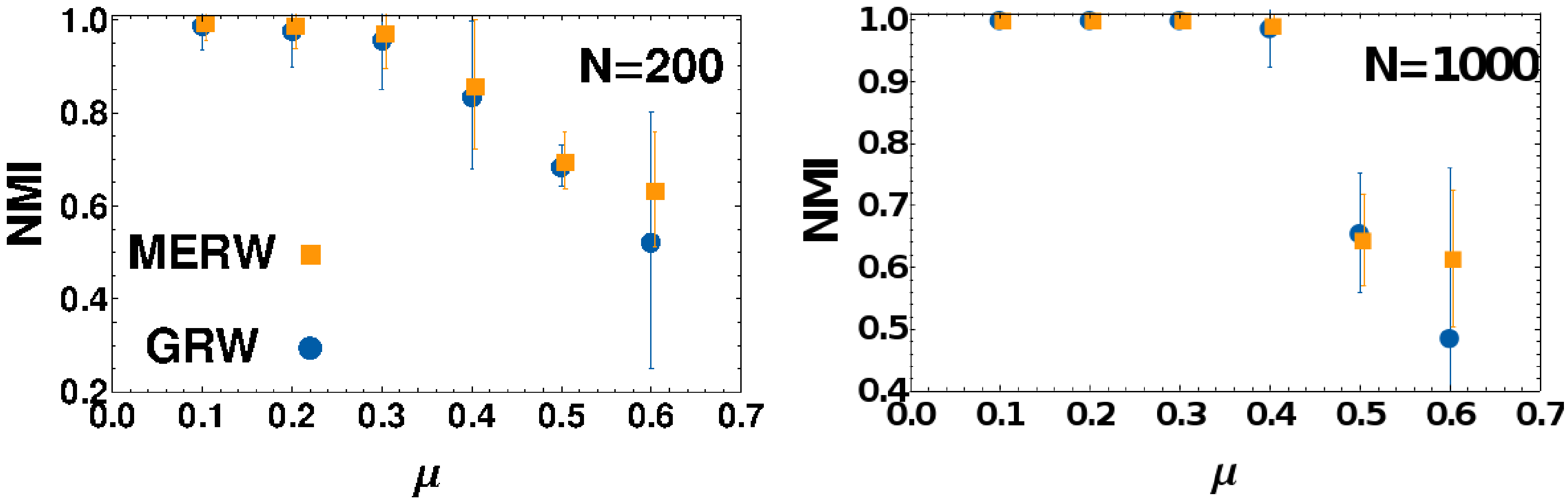
n – number of nodes in a graph

n_h – number of nodes in a comm. h

Powers of the transition matrix

$$\sum_{t=1}^{k} \mathbf{P}^{t}$$

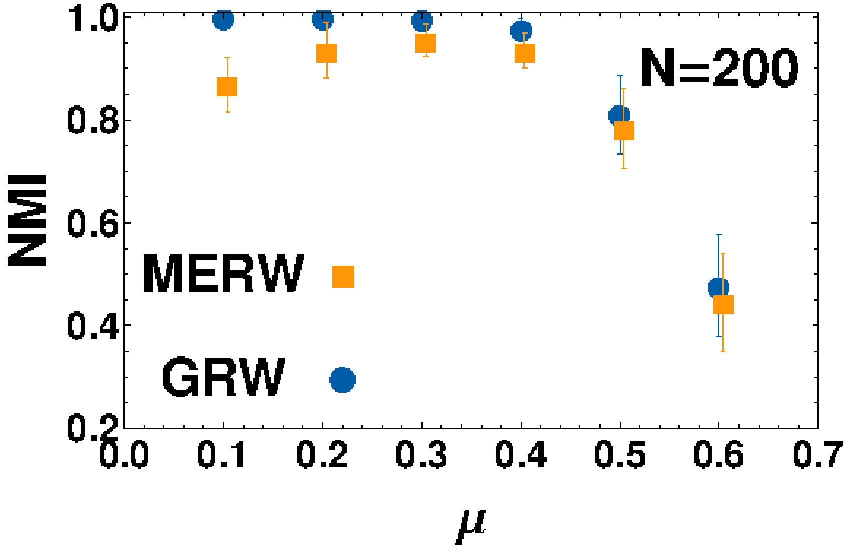
The distance between two nodes = the distance between matrix rows.
Iterated expansion of the distance.

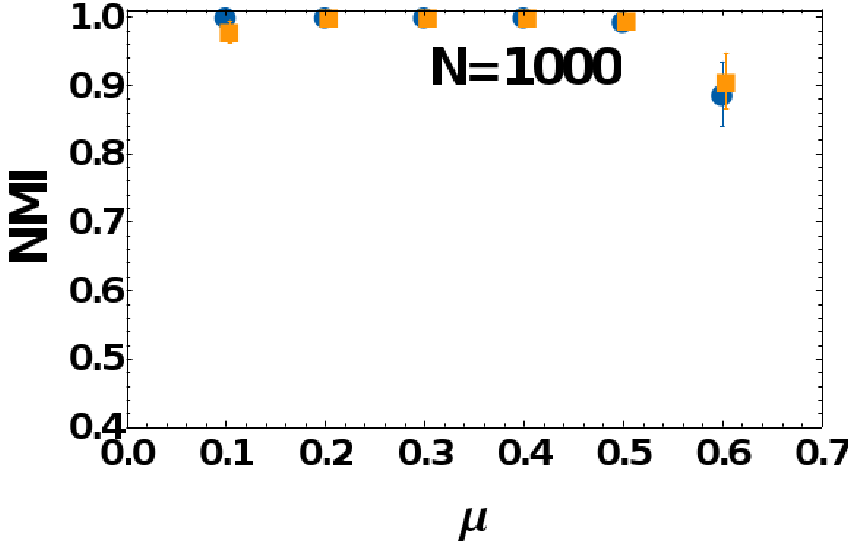


D. Harel, Y. Koren, On clustering using random walks, in: FST TCS ‘01: Proceedings of the 21st Conference on Foundations of Software Technology and Theoretical Computer Science, Springer-Verlag, London, UK, 2001, pp. 18-41

Powers of the transition matrix

The distance matrix: $r(t)_{\mathrm{ij}} = \sqrt{\sum_k \frac{\left((\boldsymbol{P}^t)_{\mathrm{ik}} - (\boldsymbol{P}^t)_{\mathrm{jk}}\right)^2}{d_k}}$





M. Latapy, P. Pons, Lect. Notes Comput. Sci. 3733 (2005) 284-293 joint with S. Fortunato

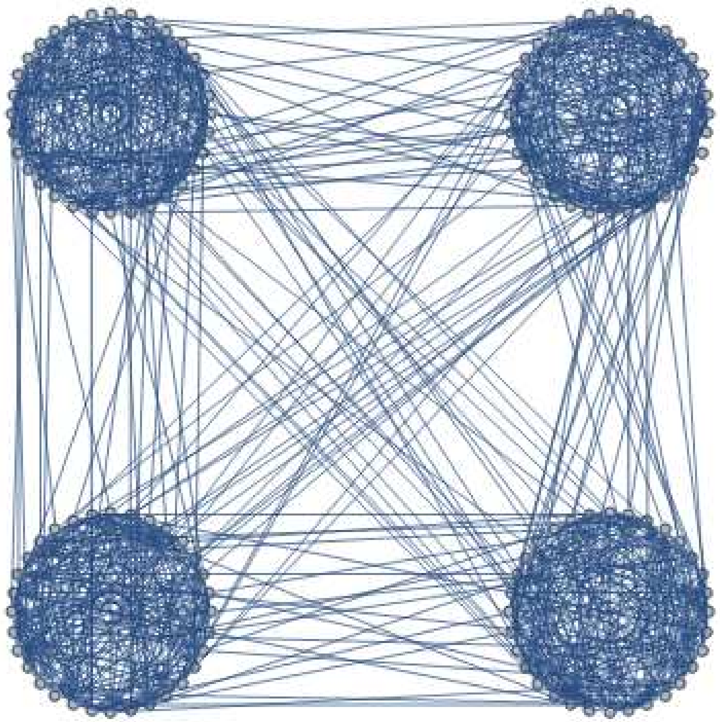
Aalto University School of Science

Espoo, Finland

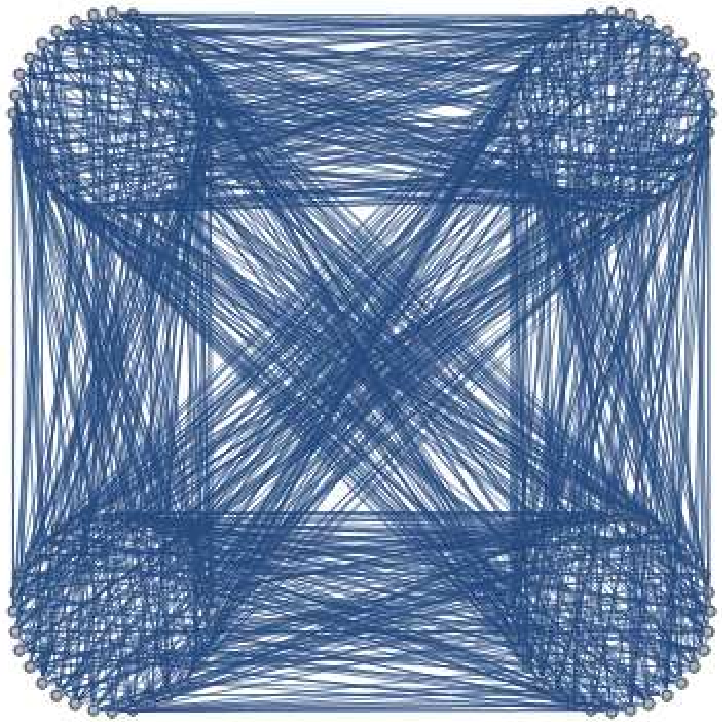
DETECTABILITY LIMITS

Girvan-Newman benchmarks

1/16 links outside



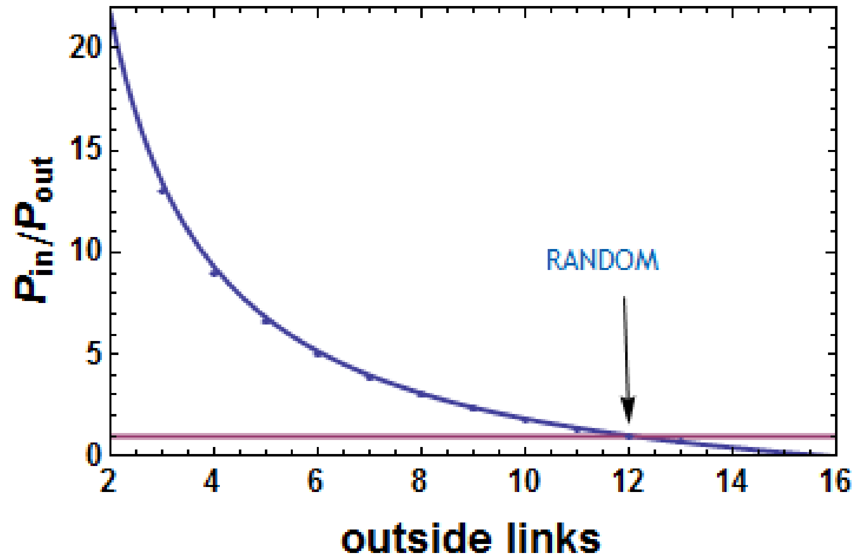
12/16 links outside

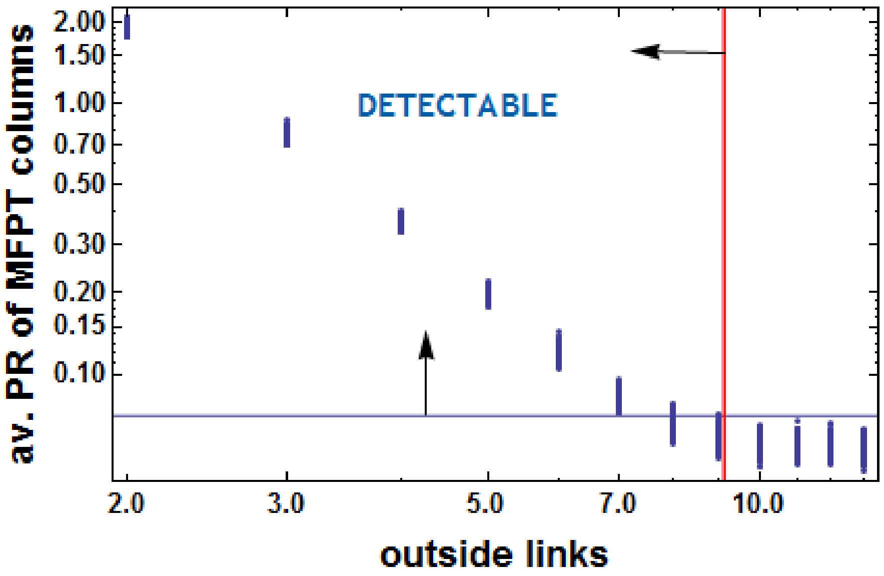


- P_{in} – prob. of in-group link
- P_{out} – prob. of outside link

M. Girvan, M.E.J. Newman, Proc. Natl. Acad. Sci. USA 99 (12) (2002) 7821

Detectability limit





P_{in} – prob. of in-group link

P_{out} – prob. of outside link

$$N - \sum_f \left(\sum_i M_{if}{}^4 \right)^{-1}$$

A. Decelle, et al., Phys. Rev. Lett. 107, 065701 (2011)

R. R. Nadakuditi and M. E. J. Newman, Phys. Rev. Lett. 108, 188701 (2012)

Conclusions

- MERW unifies a number of centrality measures
- MERW and GRW result in comparable performance of community detection methods
- There is RW indicator of detectability

Thank you

INNOVATIVE ECONOMY
NATIONAL COHESION STRATEGY

FNP
Foundation for Polish Science

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND

Project carried out within the MPD programme "Physics of Complex Systems" of the Foundation for Polish Science and cofinanced by the European Regional Development Fund in the framework of the Innovative Economy Programme.

RW-based centralities

EIGENVECTOR	$\vec{\mathbf{I}}_{\psi} = \vec{\psi}_0$
STATIONARY STATE	$\vec{\mathbf{I}}_{\pi} = \vec{\pi}$
WEIGHTED PATHS	$\vec{\mathbf{I}}_G = (\mathbf{G}(\lambda) - \mathbf{1})\vec{\mathrm{e}}$
STOCHASTIC MATRIX	$\vec{\mathbf{I}}_P = \sum_{t=1}^{T} \vec{\pi}(0)^T \mathbf{P}^t$
MFPT	$\vec{\mathbf{I}}_{MFPT} = N / \sum_i M_{if}$

$$\mathbf{G}(\lambda) = \sum_{t=0}^{\infty} \lambda^{-t} \mathbf{A}^t$$

Other similarities

$$\pi_f \sum_{t=0}^{\infty} (\mathbf{P}^t)_{fi} = \sqrt{\pi_f} G_{fi} \sqrt{\pi_i}$$

Def. similar to fundamental matrix

$$r(t)_{ij} = \sqrt{\frac{\sum_k [(\mathbf{P}^t)_{ik} - (\mathbf{P}^t)_{jk}]^2}{\pi_k}}$$

Distance matrix for a community detection alg.

$$\mathbf{r}^2(t) = \mathbf{D}[(\mathbf{P}^{2t})_{dg}\mathbf{E} - (\mathbf{P}^{2t})^T] + [\mathbf{E}(\mathbf{P}^{2t})_{dg} - \mathbf{P}^{2t}]\mathbf{D}$$

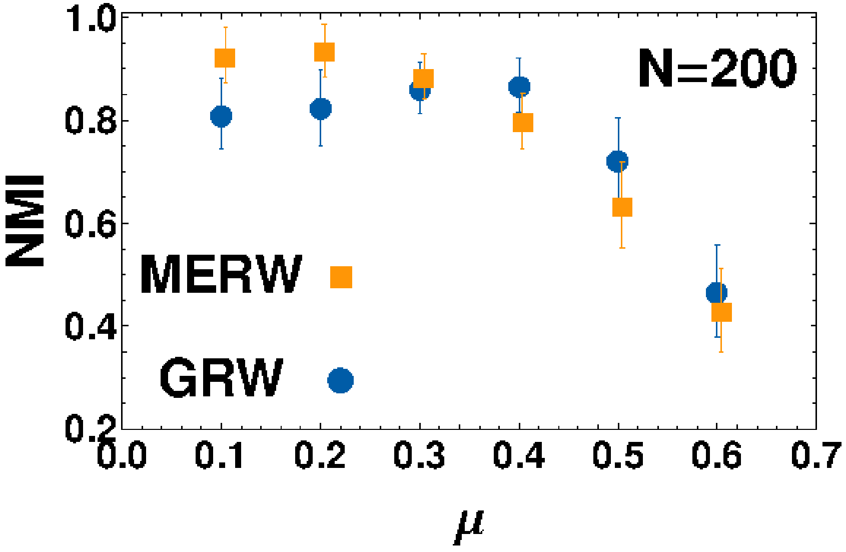
$$\mathbf{r}^2 = \mathbf{D}(\mathbf{G}^2)_{dg}\mathbf{E} - 2\sqrt{\mathbf{D}}\mathbf{G}^2\sqrt{\mathbf{D}} + \mathbf{E}(\mathbf{G}^2)_{dg}\mathbf{D}$$

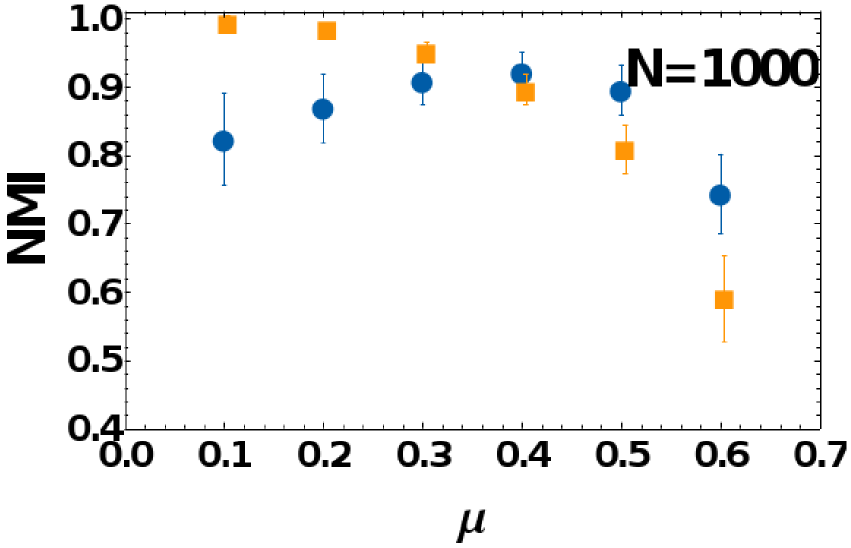
symmetric version of MFPT!

Path weighting

$K(\beta) = e^{\beta \boldsymbol{A}}$ (factorial path weighting = heat kernel)

$G = (\lambda_0 \mathbf{1} - \boldsymbol{A})^{-1}$ (exponential path weighting = resolvent = quantum particle propagator = MERW)



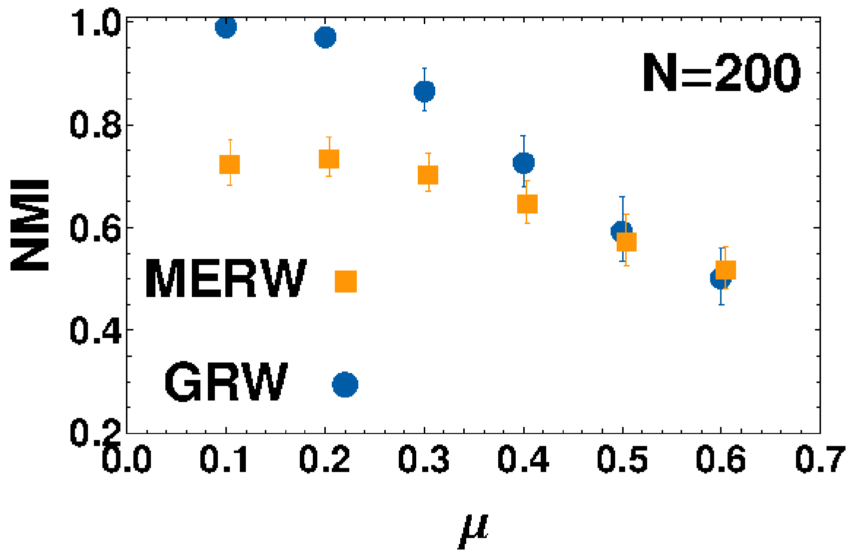


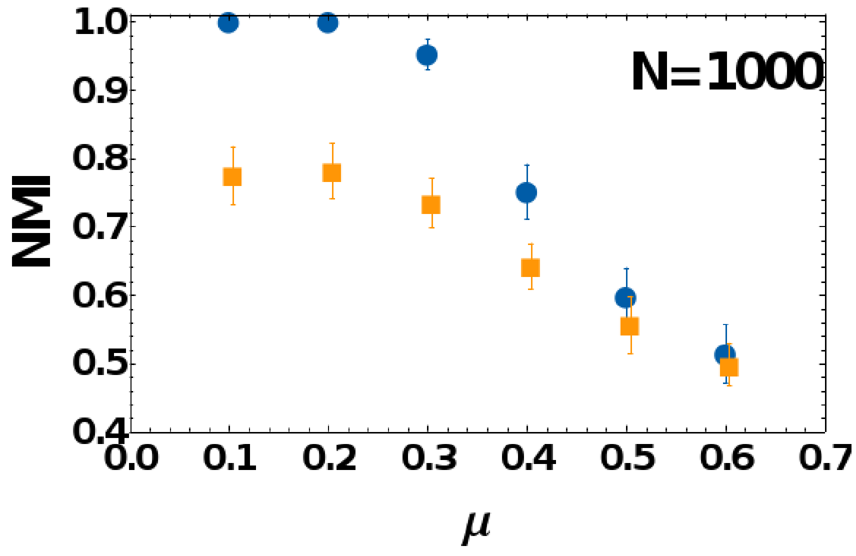
E. Estrada, N. Hatano, Phys. Rev. E 77 (3) (2008) 036111

E. Estrada, N. Hatano, Appl. Math. Comput. 214 (2009) 500-511

Mean first-passage time

The distance matrix: $$r(t)_{ij} = \sqrt{\sum_k \left(\boldsymbol{M}_{ik} - \boldsymbol{M}_{jk}\right)^2}$$





H. Zhou, R. Lipowsky, Lect. Notes Comput. Sci. 3038 (2004) 1062-1069

J.K. Ochab, Z. Burda, arXiv:1208.3688 [physics.soc-ph]

Detectability limit

As far:

- works irrespective of the type of random walk (GRW or MERW)
- seems to work for LFR benchmarks as well

To do:

- compute the random-graph limit
- check for other RWs
- check if detection methods obey it