Maximal-Entropy Random Walk
Centrality measures and communities

Jeremi K. Ochab, Z. Burda
Marian Smoluchowski Institute of Physics, Jagiellonian University
jeremi.ochab@uj.edu.pl

1. Introduction

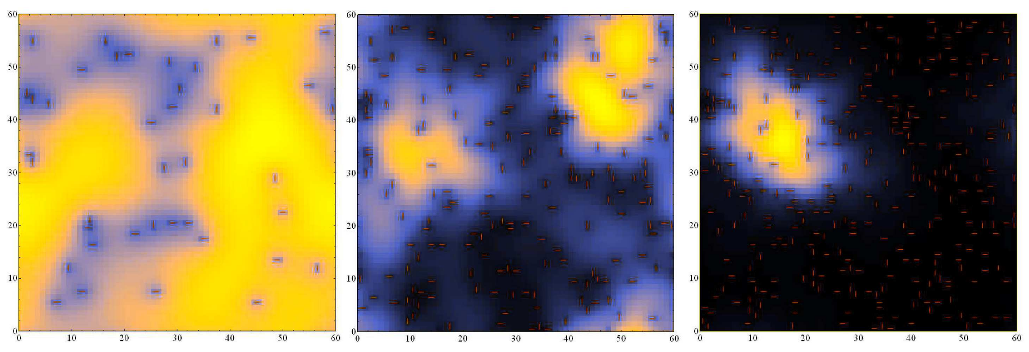

Figure 1: *Localisation of Maximal-Entropy Random Walk's stationary probability on a 2D square defected lattice.*
Interactive demonstration at: HTTP://DEMONSTRATIONS.WOLFRAM.COM/GENERICRANDOMWALKANDMAXIMALENTROPYRANDOMWALK/

GRAPHS represent abstracted relationships between entities. Upon their structure a process may take place, often formalized into the mathematical concept of random walk (RW). Together, graph and random walk can constitute a model for processes and interactions on networks.

Whatever the exact nature of the phenomena, two natural questions arise: Which entity in the network is the most influential, be it a a transcription factor, a frequented website or a renowned researcher? Do any entities form groups more strongly connected to each other than to the rest?

These two questions are answered by **centrality measures** and **community detection algorithms**.

THERE ARE TWO ANTIPODAL NULL MODELS among RWs: **Generic Random Walk (GRW)** and **Maximal-Entropy Random Walk (MERW)**. Whereas GRW relies on locally equiprobable moves (maximal entropy of the nearest neighbor selection), MERW utilises equiprobable paths (maximal entropy of the path selection). MERW provides a **unifying view of several centrality measures**. Its distinct behaviour (**localisation**, see Fig.1; **fast dynamics** on trees; **slow dynamics** between identical subgraphs) is a reason for the **comparison of community detection methods** using RWs.

2. Definitions

DISCRETE-TIME RANDOM WALK ON A GRAPH (undirected finite connected) is a Markov chain. Its unique stationary state (a probability distribution after infinite time) satisfies

$$\vec{\pi}^T \mathbf{P} = \vec{\pi}, \quad \sum_a \pi_a = 1 \,.$$

The two random walks of our interest are defined as follows:

	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. Transition matrix obeys $P_{ba} \geq 0, \sum_b P_{ba} = 1$, it is symmetric, constant with respect to time.

$P(\gamma_{a_t a_0}) = P_{a_t a_{t-1}} ... P_{a_2 a_1} P_{a_1 a_0}$ for fixed endpoints a_0, a_t is the probability of a path $(a_0, a_1, \ldots, a_t)$.

References

[1] Z. Burda, J. Duda, J.M. Luck, and B. Waclaw, *Phys. Rev. Lett.* **102**, 160602 (2009).
[2] J.K. Ochab, arXiv:1206.4094v1 [physics.soc-ph].
[3] J.K. Ochab, Z. Burda, arXiv:1208.3688v1 [physics.soc-ph].
[4] M. Latapy, P. Pons, Lect. Notes Comput. Sci. **3733**, 284-293 (2005).
[5] C. M. Grinstead, J. L. Snell, *Introduction to Probability* (American Mathematical Society, Providence, RI, 1997).
[6] D. Harel, Y. Koren, 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.
[7] E. Estrada, N. Hatano, Phys. Rev. E **77**(3), 036111 (2008); Appl. Math. Comput. **214**, 500-511 (2009).
[8] H. Zhou, R. Lipowsky, Lect. Notes Comput. Sci. **3038**, 1062-1069 (2004).

Acknowledgments

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.

INNOVATIVE ECONOMY NATIONAL COHESION STRATEGY
FNP Foundation for Polish Science
EUROPEAN UNION EUROPEAN REGIONAL DEVELOPMENT FUND

3. Analogies between (dis)similarity matrices

THE CORE QUANTITY often analysed by methods of both assessing centrality and detecting communities is a **(dis)similarity matrix**. Its entries contain the information on how strongly a pair of nodes in a network is (dis)associated.

A natural choice for a dissimilarity matrix of a graph with respect to a random walk is the **mean first-passage time** (MFPT) matrix. Its elements M_{ij} encode the average time to reach the vertex j from i for the first time (in general $M_{ij} \neq M_{ji}$). Its construction with the use of the **fundamental matrix** $\mathbf{Z}$ is given in [5]

$$\begin{aligned} \mathbf{Z} &= \left(\mathbf{1} - \mathbf{P} + \vec{\mathbf{e}}\vec{\pi}^T\right)^{-1} \\ \mathbf{M} &= (\mathbf{E}\mathbf{Z}_{dg} - \mathbf{Z})\mathbf{D} \,, \end{aligned} \tag{1}$$

where $\mathbf{1}$ is the identity matrix, $\tilde{\mathbf{e}} = (1, 1, ..., 1)^T$, $\mathbf{E}$ is a matrix of all ones, $\mathbf{Z}_{dg}$ is a diagonal matrix with elements $(\mathbf{Z}_{dg})_{ii} = Z_{ii}$, and $\mathbf{D}$ is a diagonal matrix with the diagonal elements equal to the inverses of the stationary state of a RW, $(\mathbf{D})_{ii} = 1/\pi_i$. The fundamental matrix is constructed so as to include all powers of the transition matrix: $\mathbf{1} + \mathbf{P} + \mathbf{P}^2 + \ldots$.

ANOTHER WIDELY UTILISED APPROACH is calculation of powers of the stochastic matrix $\mathbf{P}$. For instance, in [4] the dissimilarity matrix used is

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

which measures the difference between the probability distribution (a row of $\mathbf{P}^t$) as seen from the perspective of nodes i and j, and where the division by π_k is supposed to reduce the effect of a vertex's centrality. In the case of MERW and GRW (generally, for any RW for which $\mathbf{D}^{-1/2}\mathbf{P}\mathbf{D}^{1/2}$ is symmetric), it can be rewritten in the form analogous to the MFPT matrix

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

Upon including all times t, one obtains

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

where matrix $\mathbf{G}$ plays the same role as the fundamental matrix before and is defined below.

THE SIMILARITY MATRIX $\mathbf{G}$ gives the average **number of paths** between two given nodes with weights that depend on the paths' length

$$\mathbf{G}(\lambda) = \sum_{t=0}^{\infty} \lambda^{-t}\mathbf{A}^t \,. \tag{3}$$

For $\lambda > \lambda_0$ the sum is convergent and can be carried out with the use of spectral decomposition of $\mathbf{A}$.

These three intimately related quantities (1)-(3) constitute a common framework for a number of centrality measures [2, 3].

4. Centrality measures

METHODS OF MEASURING IMPORTANCE introduced during last 60 years involve concepts originating from graph theory (the degree of a vertex, enumeration of paths or the principal eigenvector of the adjacency matrix) and the theory of Markov chains (stationary states of random walks, their stochastic matrices, and mean first-passage times).

Those presented here produce significantly distinct results if defined with GRW than with MERW. Additionally, results of different methods using MERW are nearly equivalent.

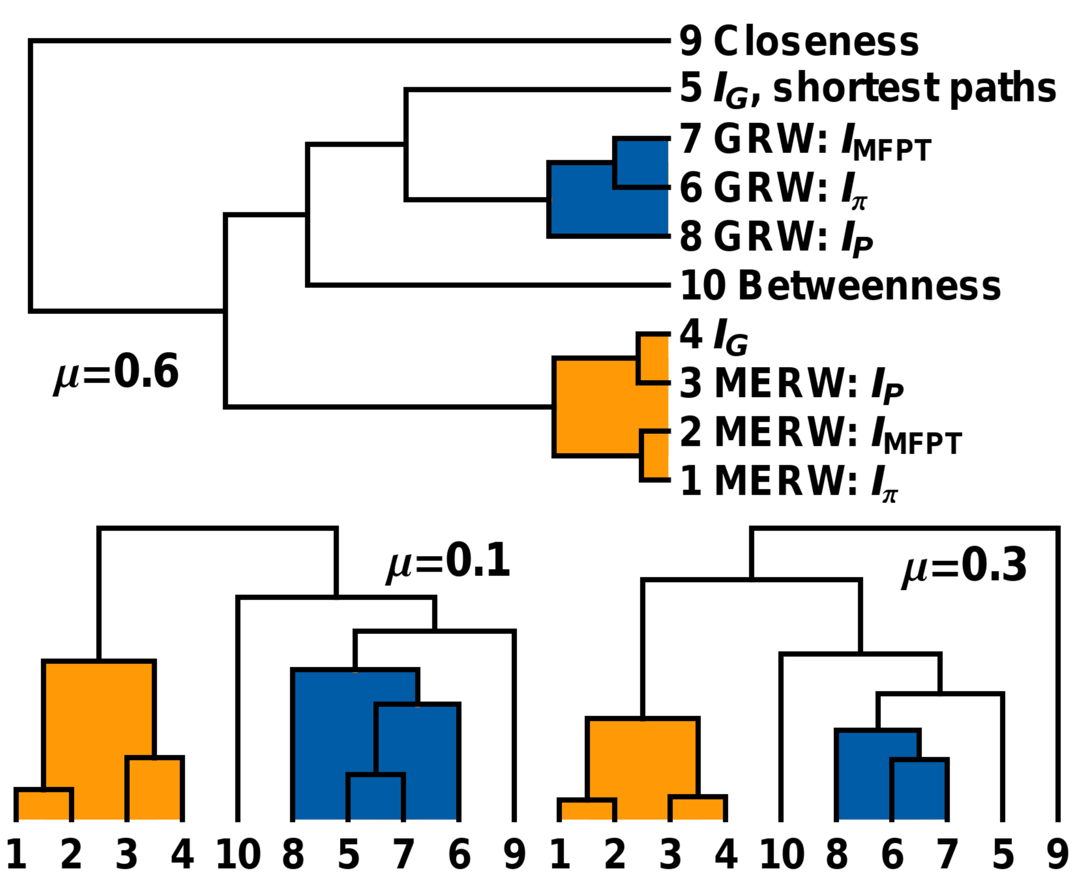

Figure 2: *The dendrograms show affinity between different centrality measures.* ***Blue*** *group corresponds to GRW but for graphs containing evident communities (small μ) it mixes with other methods.* ***Orange*** *group corresponds to measures unified by MERW; its members are on average closer to each other and further away from other measures than the members of GRW group.*

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	$\vec{\mathbf{I}}_{MFPT} = N/\sum_i M_{if}$

6. Benchmark graphs

UNWEIGHTED UNDIRECTED LFR BENCHMARK GRAPHS are used. The mixing parameter μ is the fraction of links a given node shares with the nodes outside its community. For each value μ, 100 benchmark graphs are generated. They have $N = 200, 1000$ nodes (comparison of centralities uses only $N = 200$); exponent for the degree distribution is $\tau_1 = -2$ and for the community size distribution $\tau_2 = -1$.

For $N = 200$ the parameters are: the average degree of 10, maximum degree of 30, and the minimum and maximum community sizes are taken to be 5 and 35.

For $N = 1000$: the average degree of 20, maximum degree of 50, and the minimum and maximum community sizes are 20 and 100, respectively.

5. Community detection

WHEREAS MERW exhibits surprising stationary and dynamic properties on some defective regular graphs, on the locally random LFR benchmark graphs it can offer a performance of community finding methods comparable to GRW.

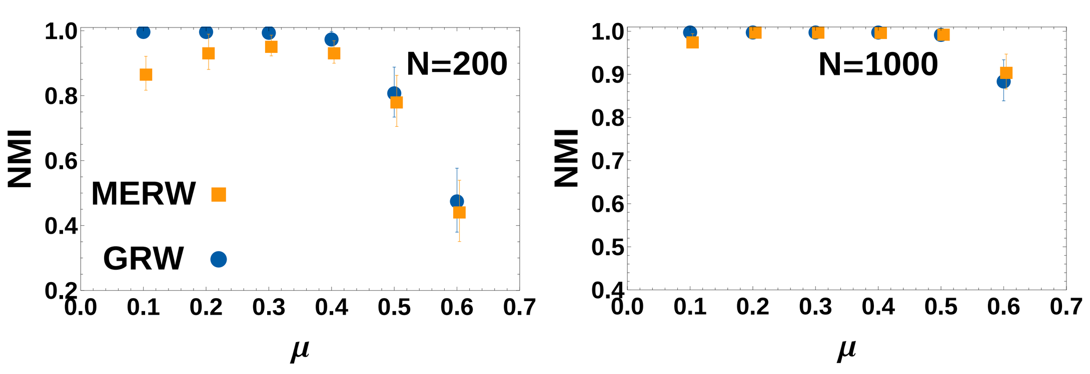

Figure 3: *Method [4] using the dissimilarity matrix* $\mathbf{r}^2(t)$ (2).

The most reliable method checked is [4] (see Fig.3), for which GRW and MERW produce similar results. Significant worsening is noted, however, for small networks when using the latter.

The other methods have not been previously compared on LFR benchmark graphs. The one in Fig.4 is not reliable for $\mu > 0.4$, although its performance is slightly improved by MERW. The one in Fig.5 used simple hierarchical clustering as temporary means. Despite this simplification, switching to MERW performs reasonably well. *Netwalk* in Fig.6 is not suited for MERW. Even for GRW, used originally, it produces an unsatisfactory results for the medium range of μ.

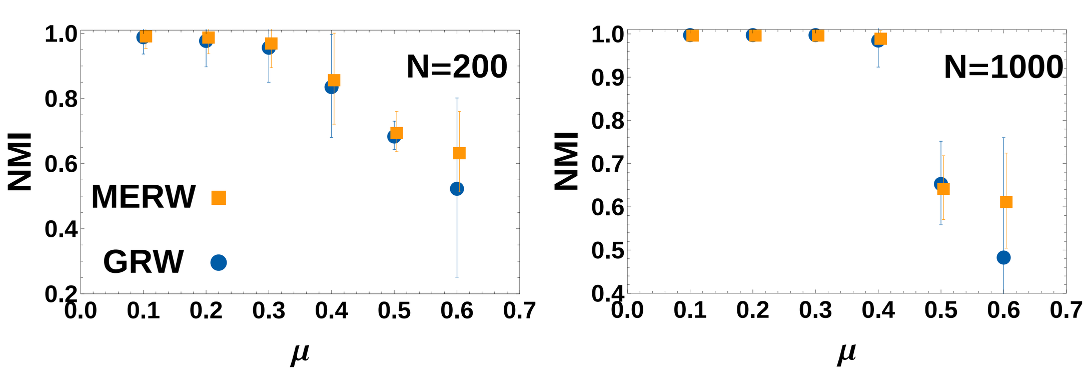

Figure 4: *Method [6] filtering powers of the transfer matrix* $\mathbf{P}$.

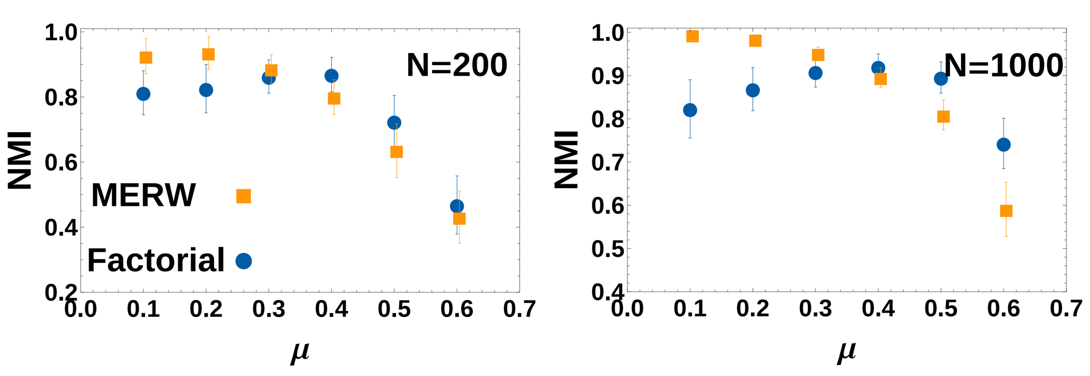

Figure 5: *Method based on matrix* $\mathbf{G}(\lambda_0)$ (3) *with exponential (MERW) or factorial [7] path weighting.*

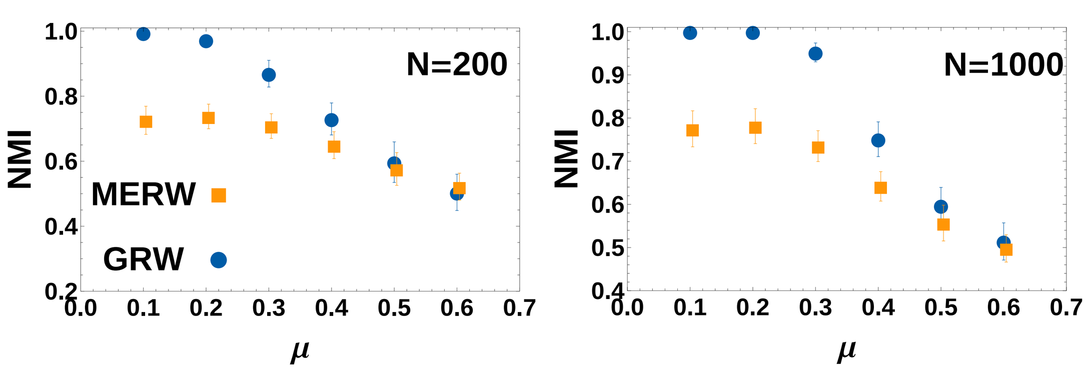

Figure 6: *Method [8] based on similarity matrix derived from MFPT matrix* $\mathbf{M}$ (1).