Maximal entropy random walk in community detection

J.K. Ochab\(^a\) and Z. Burda\(^b\)

Marian Smoluchowski Institute of Physics and Mark Kac Complex Systems Research Center
Jagiellonian University, Reymonta 4, 30-059 Kraków, Poland

Received 30 November 2012 / Received in final form 6 December 2012
Published online 31 January 2013

Abstract. The aim of this paper is to check feasibility of using the maximal-entropy random walk in algorithms finding communities in complex networks. A number of such algorithms exploit an ordinary or a biased random walk for this purpose. Their key part is a (dis)similarity matrix, according to which nodes are grouped. This study encompasses the use of a stochastic matrix of a random walk, its mean first-passage time matrix, and a matrix of weighted paths count. We briefly indicate the connection between those quantities and propose substituting the maximal-entropy random walk for the previously chosen models. This unique random walk maximises the entropy of ensembles of paths of given length and endpoints, which results in equiprobability of those paths. We compare the performance of the selected algorithms on LFR benchmark graphs. The results show that the change in performance depends very strongly on the particular algorithm, and can lead to slight improvements as well as to significant deterioration.

1 Introduction

Relationships between entities can be represented as a graph structure upon which some process takes place, be it information or opinion spread on social networks, including citation and collaboration networks, WWW or the Internet, or perhaps a physical process (molecular motion) on physical or biological networks. One of the natural questions to be asked is whether there are groups of entities which are connected stronger to each other than to the rest of the network. Due to the sociological legacy, these are called communities, but they can comprise researchers, websites, genes or transcription factors as well.

A plenitude of methods have been devised to find such communities, and a plenitude of definitions have been conceived to tell what it is that we really look for. These definitions and methods have been thoroughly reviewed in [1]. A particular subgroup of algorithms is based on random walks (RWs), since intuitively a random walker is...
expected to spend a longer time inside the well-connected graph regions, and there should be only a slim chance that it crosses from one to another.

The most common choice for such algorithms has been the well-known random walk defined by equal probabilities of going from a node to any of its nearest neighbours, which we call the generic random walk (GRW). On the contrary, maximal-entropy random walk (MERW) ensures equiprobability of all paths of a given length and endpoints. Although for many problems, GRW and biased RWs are often more suitable, MERW deserves particular interest: while the former maximises the entropy locally (entropy of the nearest neighbour selection), the latter maximises the entropy globally (entropy of the path selection) [2,3]. Among its curious behaviours, MERW exhibits localization of its stationary distribution on diluted lattices [2–4] and Cayley trees [5,6], it also relaxes extremely fast on these trees [5,7], while it does very slowly between two identical connected regions [8]. Thus, we believe MERW can serve alongside GRW as a null model of random processes on networks.

It is noteworthy that equiprobable paths (as generated by MERW) are the natural choice for an ensemble used in Feynman path integrals (e.g., discrete quantum gravity models with curved space-time) [3] or in the optimal sampling algorithm in the path-integral Monte Carlo methods [9]. Entropy maximization is a global principle much like the least action principle. It has earlier led to the biological concept of evolutionary entropy [10]. Interestingly, the value of entropy for a given graph, as defined by MERW, has been found useful for selection of robust networks [11]. Finally, it has begun to be used in the study of complex networks [12–16].

2 Generic and maximal-entropy random walks

Let us consider a discrete time random walk on a finite connected undirected graph, with its stochastic matrix \( P \) being constant in time. An element \( P_{ij} \geq 0 \) of this matrix encodes the probability that a walker that stands on a node \( i \) at time \( t \) hops to a node \( j \) at time \( t + 1 \). These matrix elements fulfil the condition \( \sum_j P_{ij} = 1 \) for all \( i \), which means that the number of walkers is conserved. An additional assumption allows the walkers to hop only to a neighbouring node. This can be formulated as \( P_{ij} \leq A_{ij} \), where \( A_{ij} \) is the corresponding element of the adjacency matrix \( A \) of the graph: \( A_{ij} = 1 \) if \( i \) and \( j \) are neighbours, and \( A_{ij} = 0 \) otherwise.

For any time \( t \), the probability of a walker staying on a given vertex of the graph is encoded in the vector \( \pi(t) = (\pi_1(t), \ldots, \pi_N(t))^T \). The initial distribution of particles is \( \pi(0) \), and the distribution after \( t \) steps \( \pi(t)^T = \pi(0)^T P^t \). A quantity of interest is the stationary probability distribution, which we assume to exist. Then, it is given by a solution of
\[
\pi^T = \pi^T P,
\]
and may be regarded as the probability distribution after infinite time.

GRW is realised by the following stochastic matrix:
\[
P_{ij} = \frac{A_{ij}}{k_i},
\]
where \( k_i = \sum_j A_{ij} \) denotes the node degree. The factor \( 1/k_i \) in the above formula produces a uniform probability of selecting one of the \( k_i \) neighbours of the node \( i \). This choice maximises the entropy of neighbour selection and corresponds to the standard Einstein–Smoluchowski–Polya random walk. The stationary probability distribution of GRW is given by \( \pi_i = k_i / \sum_j k_j \).

The other type of random walk, MERW, is defined by a stochastic matrix that maximises entropy of a set of trajectories with a given length and end-points.
This is a global principle similar to the least action principle. It leads to the following stochastic matrix:

\[ P_{ij} = \frac{A_{ij}}{\lambda_0} \psi_{0j}, \]

(3)

where \( \lambda_0 \) is the largest eigenvalue of the adjacency matrix \( A \), and \( \psi_{0i} \) is the \( i \)-th element of the corresponding eigenvector \( \psi_0 \). By virtue of the Frobenius-Perron theorem, all elements of this vector are of the same sign, because the adjacency matrix \( A \) is irreducible. For a stochastic matrix to maximise the entropy of an ensemble of paths, the choice (3) is unique.

The defining condition of entropy maximization leads to equiprobability of paths. More precisely, let us take a sequence of nodes \( \gamma_{a_0a_T} = (a_0, a_1, \ldots, a_T) \), which is a path of \( T \) steps with the initial node \( a_0 \) and the final node \( a_T \). The probability of visiting this sequence of nodes is

\[ P(\gamma_{a_0a_T}) = P_{a_0a_1} P_{a_1a_2} \cdots P_{a_{T-1}a_T}, \]

(4)

which results from the Markov property of the random walk. Upon substitution of MERW's stochastic matrix, one obtains

\[ P(\gamma_{a_0a_T}) = \frac{1}{\lambda_0^T} \psi_{0a_0} \psi_{0a_T}, \]

(5)

which depends only on the number of steps and on the two end-points, but is independent of the intermediate nodes. This is what we mean by equal probability of paths of a given length and end-points. Consequently, the probability measure on this ensemble of paths is uniform, and its entropy is maximal.

The stationary state of MERW is given by Shannon–Parry measure \[ 17\]:

\[ \pi_i = \psi_{0i}^2, \]

(6)

The last formula forms a connection between MERW and quantum mechanics, since \( \psi_{0i} \) can be understood as the wave function of the ground state of the operator \( -A \) and \( \psi_{0i}^2 \) as the probability of finding a particle in this state \[ 2,3\]. The two types of random walk, (2) and (3), behave identically on \( k \)-regular graphs. In general, however, they have completely disparate properties.

### 3 (Dis)similarity matrices for community finding algorithms

Methods of both assessing centrality \[ 18\] and finding communities \[ 19,20\] have widely utilised calculating powers of the stochastic matrix. The one by Latapy and Pons \[ 20\] uses the dissimilarity matrix

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

(7)

where the division by \( \pi_k \) is supposed to reduce the effect of centrality of a vertex. Originally, \( P \) and \( \pi \) corresponding to GRW were chosen.

Another approach is an explicit use of the mean first-passage times (MFPT) \[ 21–23\]. MFPT matrix \( M \) is a useful and well-studied quantity characterising RWs. Its construction with the use of the fundamental matrix \( Z \) is given in \[ 24,25\]

\[ Z = (1 - P + e\pi^T)^{-1}, \]

(8)

\[ M = (EZ_d - Z)D, \]

(9)
where $I$ is the identity matrix, $e = (1, 1, ..., 1)^T$, $E$ is a matrix of all ones, $Z_d$ is a diagonal matrix with elements $(Z_d)_{ii} = Z_{ii}$, and $D$ is a diagonal matrix with elements $(D)_{ii} = 1/\pi_i$. The elements $M_{ij}$ encode the average time to reach the vertex $j$ from $i$ for the first time (in general $M_{ij} \neq M_{ji}$).

The last approach we discuss is a similarity matrix containing the average number of paths between two given nodes (which is just $A^t$) with weights that depend on the length of the path.

$$G(\mu) = \sum_{t=0}^{\infty} e^{-\mu t} A^t. \quad (10)$$

For $e^\mu \equiv \lambda > \lambda_0$, the sum is convergent and can be carried out with the use of spectral decomposition of $A$. From the point of view of statistics of paths $G(\mu)$ defines the grand-canonical ensemble of paths. An element $G_{fi}(\mu)$ corresponds to the grand canonical partition function, $\mu$ corresponds to the chemical potential, and the average path length is $\langle t \rangle_{fi} = -(\ln G'_{fi}(\mu))$. To avoid a conflicting notation, henceforth we use $\lambda \equiv e^\mu$, whereas the symbol $\mu$ will be exclusively reserved for the mixing parameter of benchmark graphs (see Sect. 4.2).

In the case of MERW and GRW (generally, for any RW for which $D^{-1/2}PD^{1/2}$ is symmetric) it can be shown that these three quantities are intimately related constituting a common framework for a number of centrality measures [3, 26].

### 4 Community detection algorithms

#### 4.1 Comparison

Each of the above quantities has an analogic centrality measure: $r$ has the stationary state centrality and centralities defined by summation of powers of the stochastic matrix, $G$ has the eigenvector centrality and centralities defined by path enumeration, and $M$ has a centrality defined by the inverse of its average rows [26]. These are natural counterparts to some community finding methods.

Just as centrality may be defined with the use of the principal eigenvector of the adjacency matrix or the stochastic matrix (then, the eigenvector is the stationary state), there is a family of community finding methods analysing the rest of the eigenvectors (often it is the spectrum of Laplacian that is analysed) [27–32]. However, having the two random walks at hand, we are more interested in methods that utilize their characteristics. Particularly, we try to assess what difference it makes, when we switch between those two random walks.

Below, we present several available methods that originally use GRW as the random walk of choice. These algorithms have not been previously systematically compared on benchmark graphs (described in detail in Sect. 4.2). We measure their performance on a set of such graphs, and compare it with the performance of the same methods, in which we substituted MERW for GRW.

There are a number of methods using powers of the transition matrix. For instance, [19] use the matrix

$$P^{\leq T} \equiv \sum_{t=1}^{T} P^t, \quad (11)$$

where $P$ corresponded to GRW, and $T$ was taken around 2–3. The assumption is that two nodes are close to each other if the corresponding rows of $P^{\leq 1}$ matrix are similar.
One of the proposed similarity functions between two vectors is
\[
\text{sim}(x, y) = \exp \left( 2T - \sum_{i=1}^{N} |x_i - y_i| \right) - 1.
\] (12)

In this formula, if \( T = 1 \), the vectors \( x, y \) are rows of the stochastic matrix. Hence, the elements of each of them sum up to 1. There are \( T \) stochastic matrices summed in (11), hence in general the elements of each vector sum up to \( T \). If the two vectors are maximally different, the sum in (12) becomes \( 2T \), and the similarity reaches the lower boundary value of 0.

The algorithm consists in replacing edge weights of the original graph with the elements of the similarity matrix, so that external (intercommunity) links get smaller weights, and the internal ones get larger weights. The procedure is iterated until the differences between weights become large enough, and the weights below a given threshold can be disposed of. What remains is the communities. It is viable to use the transition matrix of MERW only in the first iteration step. As illustrated in Fig. 1, MERW produces slightly better results, especially for considerable \( \mu \). The normalized mutual information is equal to 1 when the algorithm finds the same community structure as planted in the graph, and it is equal to 0 if the two partitions are statistically independent. (Details of the benchmark graphs parameters are described in Sect. 4.2.)

Next, Pons and Latapy [20] introduced an algorithm using the quantity given in (7) as a distance matrix between nodes of the graph. Their algorithm is an agglomerative one: it starts with each node being a community, and then, based on the distance matrix, it merges the two closest adjacent communities. The condition of maximal modularity chooses the partition from the resulting dendrogram. We refer to the original paper for details.

Figure 2 shows the performance of the algorithm. For large networks, the algorithm is very good independently of the random walk chosen. For small networks, MERW considerably decreases the efficiency of the algorithm for small \( \mu \); the precise reasons for that are not established. The general tendency of the algorithms to
Fig. 2. Comparison of community detection efficiency between MERW (squares; summed powers of $P^t$, $t = 1 - 3$) and GRW (circles; $t = 3$) for the algorithm of Pons and Latapy [20]. Graph size: (a) $N = 200$, (b) $N = 1000$. It is the best among the algorithms discussed. MERW slightly decreases its performance for small $\mu$.

Fig. 3. Comparison between $\lambda_0^t$ (MERW, squares) and $t!$ (circles) path weights. Graph size: (a) $N = 200$, (b) $N = 1000$. MERW gives better performance for smaller $\mu$, while factorial weighting for larger. The overall performance is satisfactory for a method based on agglomerative clustering.

perform worse for smaller networks is probably due to small average node degree, which may result in single nodes detaching easier from their communities.

In (10), the weights $e^{-\mu t}$ produce the resolvent operator of $A$, but also factorial weights $\beta^t/t!$ might be introduced [33,34], yielding the heat kernel. To analyse the resulting matrix one needs to remove the zeroth eigenmode of $A$, so that $G$ is well-defined. The choice $e^{\mu t} = \lambda_0$ is directly related to MERW.

The procedure [33,34] goes on, producing a matrix with 0s and 1s in place of negative and positive entries of $G$. The original idea involved finding all maximal cliques (maximal complete subgraphs) of the graph represented by this matrix. Since this is computationally strenuous, we use a much simpler approach and carry out hierarchical clustering on that matrix. To obtain communities, we take the dendrogram section which maximises the modularity [35]. This algorithm, however, should be considered as only a very rough approach, just for the sake of preliminary comparison. It can be seen in Fig. 3 that exponential weights works better for small $\mu$, while factorial weights give a reasonable performance for larger values of mixing parameter.
Lastly, one may look at the methods grouping the nodes according to their MFPT values. In [21, 23], a similarity matrix is introduced that computes the total of differences between MFPTs of random walkers incoming to particular nodes $a$ and $b$ from any initial node

$$A_{ab} = \sqrt{\frac{\sum_{c \neq a, b} |M_{ac} - M_{bc}|^2}{N - 2}}.$$  

On this basis, the authors developed an algorithm called Netwalk. We skip the details of the algorithm and refer the reader to the original papers. In this case, the outcome of the comparison between MFPTs of different random walks (we also implement a biased random walk used originally by Netwalk), in Fig. 4, shows that MERW should not be used in this algorithm. The original algorithm, however, works well only for very small $\mu$, and in general its performance is unexpectedly unreliable even for large network size.

4.2 Benchmark graphs

The algorithms in Sect. 4.1 are compared to the use of unweighted undirected benchmark graphs introduced in [36] by Lancichinetti, Fortunato, and Radicchi (LFR) in a manner analogous to the authors’ later work [37]. These graphs were designed specifically to benchmark community detection methods, and they are characterized with a preset power-law distribution of node degrees, and more importantly, also with a power-law distribution of community sizes. They are constructed based on the planted partition model, in which two nodes that are a priori assigned to the same community are linked with probability $p_{in}$, and with probability $p_{out}$ if they are assigned to different communities. This means that each community is a random subgraph. The LFR benchmark graphs are parametrized in a similar manner with the mixing parameter $\mu$, which is the fraction of links that a given node shares with the nodes outside its community, and may be thought of as a fixed ratio $p_{out}/p_{in}$. The parameter $\mu$ is approximately the same for all nodes in a graph.

We take 100 benchmark graphs with $N = 200, 1000$ nodes; their exponents for the degree distribution and for the community size distribution are respectively $\tau_1 = -2$ and $\tau_2 = -1$. 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. The mixing parameter $\mu$ is set to $\mu = 0.1 - 0.6$. For the upper bound, most of the algorithms start to have severe problem with detecting communities.

To check how good partition has been found, we use the normalised mutual information (NMI) \[38\]. NMI treats node assignments to communities as probabilities. As a result, it measures the statistical independence of two assignments (probability distributions) yielding 1 if they are equivalent, and 0 if they are statistically independent. We always measure NMI of the partition obtained from a given algorithm with respect to the partition planted in the benchmark. Let us note that the definition of a community here relies on the planted partition model, which means that the performance of algorithms is checked in accordance with this particular definition.

5 Conclusions

We have briefly introduced the concept of maximal-entropy random walk and reviewed some of its features, while in the main body of this paper we compared the performance of several community finding algorithm, in which MERW-based (dis)similarity matrices substituted the original ones.

The results obtained by the most reliable method checked here, made by Latapy and Pons, are comparable for GRW and MERW, although we note a significant worsening for small networks when using the latter.

The other methods have not been previously compared on LFR benchmark graphs. The one by Harel and Koren is generally unreliable for $\mu > 0.4$. However, its performance is slightly improved by MERW for both small and large networks. By contrast, MERW does not suit for Netwalk. Even for GRW, which was used originally, this algorithm produces a markedly unsatisfactory results for the medium range of the mixing parameter in comparison with the available state-of-the-art methods. The method based on factorial path weighting has considerable problems for small $\mu$. Surprisingly, switching to exponential weighting, which corresponds to MERW, produces better results than Netwalk. In general, it performs reasonably well, even though the algorithm used a simple hierarchical clustering as a temporary means for the sake of comparison.

Meanwhile, MERW exhibits a surprising localisation and relaxation properties on some defective regular graphs. This case study shows that on the LFR benchmark graphs, which are locally random, this random walk can offer a performance of community finding methods comparable to that of GRW. It remains to be investigated if the behaviour of MERW on other types of graphs, including real-world networks, is more distinctive. Further effort is also needed to determine whether the development of a dedicated algorithm which makes better use of the information contained in this type of random walk is possible.

Project operated within the Foundation for Polish Science International Ph.D. Projects Programme co-financed by the European Regional Development Fund covering, under the agreement no. MPD/2009/6, the Jagiellonian University International Ph.D. Studies in Physics of Complex Systems.

References

1. S. Fortunato, Phys. Rep. 486, 75 (2010)
2. Z. Burda, J. Duda, J.M. Luck, B. Waclaw, Phys. Rev. Lett. 102, 160602 (2009)
3. Z. Burda, J. Duda, J.M. Luck, B. Waclaw, Acta Phys. Pol. B 41, 949 (2010)
4. B. Waclaw, *Generic Random Walk and Maximal Entropy Random Walk*, Wolfram Demonstration Project
5. J.K. Ochab, Z. Burda, Phys. Rev. E 85, 021145 (2012)
6. J.K. Ochab, *Stationary States of Maximal Entropy Random Walk and Generic Random Walk on Cayley trees*, Wolfram Demonstration Project
7. J.K. Ochab, *Dynamics of Maximal Entropy Random Walk and Generic Random Walk on Cayley trees*, Wolfram Demonstration Project
8. J.K. Ochab, Acta Phys. Pol. B 43, 1143 (2012)
9. J.H. Hetherington, Phys. Rev. A 30, 2713 (1984)
10. L. Demetrius, V.M. Gundlach, G. Ochs, Theor. Popul. Biol. 65, 211 (2004)
11. L. Demetrius, T. Manke, Phys. A 346, 682 (2005)
12. V. Zlatic, A. Gabrielli, G. Caldarelli, Phys. Rev. E 82, 066109 (2010)
13. J.-C. Delvenne, A.-S. Libert, Phys. Rev. E 83, 046117 (2011)
14. R. Sinatra, J. Góméz-Gardeñes, R. Lambiotte, V. Nicosia, V. Latora, Phys. Rev. E 83, 030103 (2011)
15. C. Monthus, T. Garel, J. Phys. A: Math. Theor. 44, 085001 (2011)
16. K. Anand, G. Bianconi, S. Severini, Phys. Rev. E 83, 036109 (2011)
17. W. Parry, Trans. Amer. Math. Soc. 112, 55 (1964)
18. S. White, P. Smyth, *KDD ’03: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, 2003* (ACM, New York, USA, 2003), p. 266
19. D. Harel, Y. Koren, *FST TCS ’01: Proceedings of the 21st Conference on Foundations of Software Technology and Theoretical Computer Science* (Springer-Verlag, London, UK, 2001), p. 18
20. M. Latapy, P. Pons, Lect. Notes Comput. Sci. 3733, 284 (2005)
21. H. Zhou, Phys. Rev. E 67, 061901 (2003)
22. H. Zhou, Phys. Rev. E 67, 041908 (2003)
23. H. Zhou, R. Lipowsky, Lect. Notes Comput. Sci. 3038, 1062 (2004)
24. J.G. Kemeny, J.L. Snell, *Finite Markov Chains* (Springer Verlag, New York, 1976)
25. C.M. Grinstead, J.L. Snell, *Introduction to Probability* (American Mathematical Society, Providence, RI, 1997)
26. J.K. Ochab, Phys. Rev. E 86, 066109 (2012)
27. L. Donetti, M.A. Munoz, J. Stat. Mech., P10012 (2004)
28. K.A. Eriksen, I. Simonsen, S. Maslov, K. Sneppen, Phys. Rev. Lett. 90, 148701 (2003)
29. I. Simonsen, Phys. A 357, 317 (2005)
30. J. Shi, J. Malik, IEEE Trans. Pattern Anal. Mach. Intell. 22, 888 (2000)
31. M. Meila, J. Shi, *AI and STATISTICS (AISTATS)* (2001)
32. A. Capocci, V.D.P. Servedio, G. Caldarelli, F. Colaiori, Phys. A 352, 669 (2005)
33. E. Estrada, N. Hatano, Phys. Rev. E 77, 036111 (2008)
34. E. Estrada, N. Hatano, Appl. Math. Comput. 214, 500 (2009)
35. M.E.J. Newman, M. Girvan, Phys. Rev. E 69, 026113 (2004)
36. A. Lancichinetti, S. Fortunato, F. Radicchi, Phys. Rev. E 78, 046110 (2008)
37. A. Lancichinetti, S. Fortunato, Phys. Rev. E 80, 056117 (2009)
38. L. Danon, A. Díaz-Guilera, J. Duch, A. Arenas, J. Stat. Mech.: Theory Exp., P09008 (2005)