Spectral statistics of random geometric graphs

Carl P. Dettmann\textsuperscript{1}, Orestis Georgiou\textsuperscript{2}, and Georgie Knight\textsuperscript{1}

\textsuperscript{1} School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK, carl.dettmann@bristol.ac.uk, g.knight@bristol.ac.uk
WWW home page: \url{www.eng.ox.ac.uk/sen}

\textsuperscript{2} Toshiba Telecommunications Research Laboratory, 32 Queens Square, Bristol BS1 4ND, UK, orestis.georgiou@toshiba-trel.com

\textbf{1 Introduction}

Random geometric graphs (RGGs) were introduced to model wireless networks \cite{3}. They have subsequently found use studying other spatial networks \cite{1}, \cite{9}. In a RGG, \(N\) nodes are distributed uniformly at random throughout a given domain and are connected when within a given distance \(r\) of each other. Here our domain is the unit square with periodic boundary conditions. We study the adjacency matrix \(A\) which has entries \(a_{ij} = 1\) when there is a connection between nodes \(i\) and \(j\) and zero otherwise. The spectrum of \(A\) consists of \(N\) real eigenvalues \(\lambda_i, i = 1, \ldots, N\), and encodes information about the underlying topology and dynamics on the network \cite{8}. We study two values of \(r\) with \(N = 10^3; r = 0.09375\) which is relatively low but still ensures that the RGGs in our ensembles are connected and \(r = 0.3\) which is relatively high without introducing degeneracy into the spectrum.

Random matrix theory (RMT) has been used to study the spectrum of many real-world complex networks and random graph models \cite{5}, \cite{6}, \cite{7}. The statistics of the spacings for a correlated sequence is close to those found in the spectrum of Gaussian orthogonal ensemble (GOE) matrices from RMT. For uncorrelated sequences the spacings follow Poisson statistics. Results interpolating between the two are used to explain levels of randomness, community structure and localisation of eigenvectors. Further applications include differentiating between cancerous and healthy protein networks \cite{10}. Whilst RMT has been used to analyse many types of complex network, it has not been applied to spatial networks. Hence we apply it here to RGGs. We study short-range correlations via the nearest neighbour spacing distribution (NNSD) and long-range via the spectral rigidity \(\Delta_3\).

\textbf{2 Results}

As in RMT, in order to study the spacing statistics we need to \textit{unfold} the spectrum to a constant level density. For this we use the cumulative spectral function \(\eta(E)\) which counts how many eigenvalues there are less than or equal to a given value \(E\),

\[
\eta(E) = \sum_{i=1}^{N} \Theta(E - \lambda_i).
\] (1)
Fig. 1: Here the NNSD $P(s)$ is calculated from an ensemble of 2250, $10^3$ node RGGs with connection radius $0.9375$ (blue dots) and $0.3$ (red crosses) we compare with GOE statistics (black line). Note we have removed a peak at zero caused by the appearance of repeated eigenvalues found in RGGs.

Where $\Theta(x)$ is the Heaviside function. The unfolded eigenvalues are defined using the cumulative mean spectral function $\bar{\lambda}_i = \langle \eta(\lambda_i) \rangle$. Where $\langle ... \rangle$ is an ensemble average.

The nearest neighbour spacings $s_i = \lambda_{i+1} - \lambda_i$ and their distribution $P(s)$ we calculate from an ensemble of RGGs (Fig.1). We see $P(s)$ is close to the GOE statistics.

The spectral rigidity $\Delta_3(L)$ quantifies long range correlations in the spectrum. $\Delta_3(L, x)$ measures the least-square deviation of the unfolded cumulative spectral function $\overline{\eta}(E)$ to the line of best fit over $[x, x+L]$. Where $\overline{\eta}(E)$ counts how many unfolded eigenvalues there are less than or equal to $E$. The average over non-intersecting intervals of length $L$ is the spectral rigidity $\Delta_3(L)$. For a sequence of equal spacings (‘picket fence’ spectrum) $\Delta_3(L) = \frac{1}{12}$ whilst for Poisson spacing statistics $\Delta_3(L) = \frac{L}{15}$. For GOE there is logarithmic dependence on $L$. For large $L$ to order $1/L^2$:

$$\Delta_3(L) \simeq \frac{1}{\pi^2} \left( \ln(2\pi L) + \gamma - \frac{5}{4} - \frac{\pi^2}{8} \right).$$ (2)

Illustrated in Fig.2 is $\Delta_3(L)$ for our RGGs. As is found in other random graph models, RGGs follow GOE statistics up to some value $L_0$, then deviate. The value of $L_0$ has been related to the amount of community structure within the network [4]. In [5], $L_0/N$ is interpreted as a measure of the amount of randomness in the connections of the network. This amount of randomness is defined in terms of the rewiring probability of regular networks.

**Summary.** We numerically investigated RGGs, a model of spatial networks using random matrix theory and showed how they follow the universality found in non-spatial random graphs. Short range correlations in the spacing statistics of the spectrum were found to follow the GOE from RMT whilst for long range the connection radius was.
found to be related to deviations from GOE, a phenomenon related to the amount of randomness in non-spatial network models. Future work will investigate more general connection functions with respect to the $\Delta t$ statistic and the value of $L_0$.

We would like to thank Prof. J. P. Coon, Prof. S. Jalan and Prof. J.P. Keating for helpful discussions and comments. This work was supported by the EPSRC grant number EP/N002458/1 for the project Spatially Embedded Networks.

References

Weighted Complex Modular Hierarchy Model

Keith Smith¹,² and Javier Escudero¹

¹ Institute for Digital Communications, School of Engineering, University of Edinburgh, West Mains Rd, Edinburgh, EH9 3FB, UK, k.smith@ed.ac.uk.
² Alzheimer Scotland Dementia Research Centre, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK

1 Introduction

Looking at network models from a weighted perspective brings to light the flexibility of network simulation by reducing the problem of distributed edge probabilities to a manipulation of edge weights. Generalisations can thus be made to, e.g., nodes or modules in a straightforward manner. Smith & Escudero noted this method in their generation of the undirected Weighted Complex Hierarchy (WCH) Model [8] and showed that it emulated EEG functional connectivity networks more accurately and flexibly than classical Watts-Strogatz and Albert-Barabasi models. However, the important concept of modularity [5], where specific communities of nodes maintain stronger connections within communities than between communities, was not accounted for in this original model which was evident in comparison with modularity metric values of real EEG functional networks. In this study we present an algorithm which updates the WCH model to account for the appearance of these communities in real-world networks, thus overcoming previous limitations of the model. We show that this update indeed allows a better approximation of EEG functional networks obtained from the weighted phase-lag index (wpli). We compare our model with Random Modular Graphs [1].

2 Weighted Complex Modular Hierarchy Model

The basis of our model is the E-R randomly weighted network. Instead of generating a random number in the interval [0, 1] for considering the probability, p, of the existence of the edge, as in the sparse binary model, we instead realise the generated number as the weight of the edge. The sparse binary model can be recovered from this by selecting a binarising threshold, T = p, so that edges only exist if they exceed the threshold. The hierarchical component of the model is considered by choosing a number of hierarchy levels to which nodes are assigned in accordance with a probability space, defined on [0, 1], with respect to a suitable cumulative distribution function (cdf). Typically, this is the geometric distribution where the higher hierarchy is associated with the larger values in the cdf (least likely solutions). The other parameter of the model is the additional weight associated to nodes in given hierarchical levels, $s_{\text{hier}}$. These additional weights are then added to the weights of all the node’s adjacent edges. In the probability sense, this is equivalent to adding greater probability of existence of those edges of nodes in higher hierarchical levels [8].
The modular component is included by an additional three parameters: the additional weight associated with modules, $s_{\text{mod}}$, and the maximum, $M$, and minimum, $m$, size of modules. The algorithm then chooses $N$ random numbers $m < X_1, X_2, \ldots, X_N < M$ until $\sum_i X_i \geq n - m$ for a network of size $n$. If $\sum_i X_i < n$, the remainder, $n - \sum_i X_i$, is added to the minimal element of $\{X_i\}_{i=1}^N$. If $\sum_i X_i > n$, $\sum_i X_i$ is subtracted from the maximal element of $\{X_i\}_{i=1}^N$ ensuring that $\sum_i X_i = n$. It would be a rare exception when one of the module sizes violates the minimum-maximum settings: simulations suggest 0.05% of maximal violations, 0% minimal violations. The sizes of modules are then aligned with the model’s nodes in the standard, straightforward manner: we write the labels of the vertex set, $V$, as $\{1, \ldots, X_1, X_1 + 1, \ldots, X_1 + X_2, \ldots, X_1 + X_2 + \cdots + X_N\}$, and take module $\mathcal{V}_1 = \{1, X_1\}, \mathcal{V}_2 = \{X_1 + 1, \ldots, X_1 + X_2\}, \ldots, \mathcal{V}_N = \{X_1 + X_2 + \cdots + X_{N-1} + 1, \ldots, X_1 + X_2 + \cdots + X_N\}$. Then $s_{\text{mod}}$ is added to all weights between nodes in the same module.

### 3 Datasets

**EEG:** 64-channel EEG recordings of 109 subjects in an eyes-open resting state condition [7][2]. Connectivity matrices, averaged over trials, were computed in the Beta band (12.5-32Hz) using wpli, see [1] for further details.

**Weighted Complex Modular Hierarchy:** 109 64-node networks with $s_{\text{hier}} = 0.25$ (found to be most relevant in this case), $s_{\text{mod}} = \{0, 0.05, 0.1, \ldots, 0.7\}$, $m = 6$, $M = 19$ (10-30% of $n = 64$).

**Random Modular Graph:** Algorithm obtained online [1]. 50 iterations of each mix of parameter settings were computed for: number of modules, $\{3, \ldots, 10\}$, probability of attachment, $\{0.1, 0.2, \ldots, 1\}$, proportion of edges within communities $\{0.1, 0.2, \ldots, 1\}$.

**Metrics:** We compute and compare modularity $Q$, degree variance, $V$, and clustering coefficient, $C$, for each network.

### 4 Results and Discussion

We plot the three metrics, $Q$, $V$ and $C$, against connection density for each model and the EEG wpli networks as seen in Fig. 1. As expected, the modularity of our model increases with increasing $s_{\text{mod}}$ at all densities. This better approaches the modularity of the EEG functional connectivity (Fig. 1 left). There is an obvious inverse dependency between $Q$ and $V$, therefore a higher initial parameter of hierarchy strength is required (0.25) than that reported in [8] (0.2) to better approximate the EEG values. On the other hand, with increasing $Q$, $C$ values move away from the EEG wpli network values. The random modular graph topology has much lower $V$ and very peculiar behaviour with $C$ which make it unuseful in comparisons with the EEG network.

### 5 Conclusion

The inclusion of the module strength parameter shows a correctional influence on the previous model for approximating the modularity of real EEG networks. Our model is
Fig. 1: Modularity (Q), left, degree variance (V), centre, and clustering coefficient (C), right, against connection density (%) for the WCMH model (dark to light blue lines indicates increasing module strength parameter), EEG networks (pale yellow gives standard deviation over subjects) and random modular graph model (purple dots are the mean of one of the mixed parameter realisations).

substantially more relevant than the random modular graph model for networks with hierarchical structures. Future work will extend this work to consider possible correctional parameters for integration to make a comprehensive and highly flexible model for characterising the main aspects of network topology. We will also seek to compare our model with a wider variety of models to understand whether or not our model provides a clear advantage to those already existing [3][4][6].

References

Performance optimization of networks results in hierarchical structures with spin-glass like energy landscape

Maryam Zamani\(^1\) and Tamas Vicsek\(^1\)

Department of Biological Physics, Eotvos University, Pazmany Peter setany I/A 1117, Budapest, Hungary, zamani@caesar.elte.hu,

WWW home page: https://sites.google.com/site/maryamzamaniphysics/

Abstract. We consider complex networked organizations made of interacting units such that these interactions are favourable for both the individual units and the whole system. A simple and still realistic expression is introduced to describe the efficiency of the organization satisfying the above conditions. By updating the initially random network towards higher efficiency, locally optimal states with a hierarchical structure are obtained.

1 Introduction

Following a "Darwinian approach" (applicable to almost all existing systems, otherwise due to competition they would have already "died out") we assume that complex organizations are optimal from the point of their function. Hierarchical networks associated with many complex systems are widely observed in nature and society so we address their relation to performance. In Ref. 3 a realistic approach was developed for the spontaneous emergence of complex hierarchical networks of agents. In this paper, we describe simple rules to model how complex networks designed to perform a given general task are evolving. An efficiency function is defined to evaluate the performance based on the contribution of agents in a collaborating group. The local extremal points (maxima) of the efficiency function are obtained by using Monte Carlo simulation. These extremal points represent the locally highest performance of the networks and are achieved by optimizing the direction of the influence between the collaborating agents. As it turns out, there are many local maxima and the structure of the optimal networks is hierarchical.

2 Method

We start with Erdos-Renyi network, where each node in the graph represents a unit in a group with a certain ability between 0 and 1. The ability of individuals \(A_i\) is allowed to follow a system-dependent distribution such as bounded log normal, normal, Pareto, etc. Consider agents in a group (organization, university, etc). Agents compete to acquire more information and knowledge by interacting with other agents to maximize their
The interaction coefficient \( I_{ij} \) is +1 with probability \( \rho = (1 - \lambda)|A_i - A_j| + \lambda \), where \( \lambda \) is a constant if the direction of the edge is from higher to lower ability and is -1 (opposite direction) otherwise. An agent has a limited potential to interact with the others so for most of the \( i,j \) pairs \( I_{ij} = 0 \). Obviously, this expression needs some further interpretation. We assume that in most of the cases the interaction increases the efficiency in a way proportional to the abilities of both of the agents, while, to be realistic, we also consider that in some cases the relation of the agents is antagonistic \( (I_{ij} = -1) \) and decrease the performance of the system as a whole. Higher \( \lambda \) values decrease the probability of antagonistic relations. The local optimal states of the complex network are found by maximizing the efficiency function using Monte Carlo simulations and random relocation of randomly selected edges.
3 Results

The distribution (PDF) of local optimal states and the corresponding global reaching centrality (GRC) \[2\] (the latter measuring the level of hierarchy) have been calculated for 24 agents with bounded log-normal distribution of abilities with upper limit of 1 with \(\log \text{Mean} = -0.836, \log \text{Sdev} = 0.536\), 48 edges and various values of \(\lambda\). Results are presented in Fig. 1. 500 local maxima have been used for constructing the PDF-s. The efficiency and GRC values change significantly across the optimal states. By increasing \(\lambda\) both the efficiency and GRC values grow. \(\lambda = 0.8\) is a kind of critical value at which both phases hierarchy and lack of hierarchy, corresponding to high and low efficiency, coexist. The overall trend is that the states with higher efficiency have higher GRC values. In other words, states with higher performance are more hierarchical.

4 Conclusion

The efficiency landscapes we have obtained has analogies with those calculated for a physical model of complexity: the so called Ising spin-glass \[1\]. Antagonistic interactions - common in organizations - result in frustrations quite like in spin glasses. Since the process we define does not satisfy time reversal and has elements of randomness, we obtain a spectrum of optima even for no antagonistic relations. The results of this study can be useful when optimizing complex organizations and other kinds of systems that can be associated with networks. Larger system sizes are also attainable but the running times grow as a high power of the number of units.

References

Understanding Multi-scale Community Detection by Markov Stability with Vector Partitioning

Zijing Liu and Mauricio Barahona
Department of Mathematics, Imperial College London, London, UK,
zijing.liu@imperial.ac.uk
m.barahona@imperial.ac.uk

1 Introduction
Community detection in complex networks has attracted much attention due to its potential for practical applications, and its close mathematical relation to other problems such as data clustering, graph partitioning, and image segmentation. Markov Stability is a general framework for community detection in directed weighted networks, which employs a dynamical process on the graph to unfold the multi-scale structure of graphs [2]. In this work, we reformulate Markov Stability as a max-sum vector partitioning problem [1], and provide a geometric interpretation to elucidate the role that the Markov time plays as resolution parameter in community detection.

2 Markov Stability and Vector Partitioning
Let us consider a weighted undirected network with n nodes with associated adjacency matrix $A$, where $A_{ij} = A_{ji} > 0$ is the weight of the edge connecting nodes $i$ and $j$. The degree of the nodes is compiled as an $n$-dimensional vector $d$ with components $d_i = \sum_{j=1}^{n} A_{ij}$, and we also define the degree matrix $D = \text{diag}(d)$. The total weight of the network is $m = \sum_{i,j} A_{ij} / 2$.

Let us define a continuous-time Markov process taking place on the network and governed by the dynamics $dp/dt = -p(I - M)$, where $p$ is an $n \times 1$ row vector defined on the nodes and $M = D^{-1}A$, and the process has a unique stationary distribution $\pi = d^T / 2m$. The time of this process is denoted henceforth as the Markov time. Given a partition $g$ of the nodes into $c$ non-overlapping groups (or communities) denoted by $g = \{g_1, g_2, ..., g_c\}$, the Markov Stability of the partition is given as:

$$r(t, g) = \sum_{s=1}^{c} \sum_{i,j \in g_s} B(t)_{ij}.$$  \hfill (1)

The following proposition follows directly from the definitions.

**Proposition 1 (Spectral properties relating $B(t)$ and $M$).** Let us denote the eigenvalues of $M$ as $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k \geq ... \geq \lambda_n$ with corresponding eigenvectors $v_k = [v_{k,1}, v_{k,2}, ..., v_{k,n}]^T$, such that $Mv_k = \lambda_k v_k$ and $v_k^T \Pi v_l = \delta_{kl}$ where $\delta_{kl}$ is the Kronecker delta. Then the generalized eigenvalue problem $B(t)v = \lambda \Pi v$ is solved by the
eigenvectors \( \{1, v_2, ..., v_n\} \) with corresponding eigenvalues \( \{0, \lambda_2(t), ..., \lambda_n(t)\} \), where 
\[ \lambda_k(t) = \exp(-t(1 - \lambda_k)) > 0. \]

The following Gram matrix decomposition then follows directly:
\[ B(t)_{ij} = \lambda_2(t)\pi_i v_{2,i} \pi_j v_{2,j} + \lambda_3(t)\pi_i v_{3,i} \pi_j v_{3,j} + \ldots + \lambda_n(t)\pi_i v_{n,i} \pi_j v_{n,j} := x_i(t)^T x_j(t), \]
where \( x_i(t) := [\sqrt{\lambda_2(t)}\pi_i v_{2,i}, \sqrt{\lambda_3(t)}\pi_i v_{3,i}, \ldots, \sqrt{\lambda_n(t)}\pi_i v_{n,i}]^T, i = 1, \ldots, n. \) Hence the Markov Stability (1) of partition \( g \) can be written as:
\[ r(t, g) = \sum_{s=1}^c \sum_{i,j \in g_s} x_i(t)^T x_j(t) = \sum_{s=1}^c \left\| \sum_{i \in g_s} x_i(t) \right\|^2, \]
which is a sum of \( c \) terms, each the squared length of the sum of the vectors \( x_i \) in each community \( g_s \).

Therefore the partition that maximizes Markov Stability is equivalent to a maximum length vector partitioning problem in terms of the \( x_i \), which correspond to time-dependent eigenvectors of the generator weighted by the square root of (time-varying) eigenvalues and the degree of node \( i \). This rewriting allows us to investigate a geometric approach to community detection in the associated spectral vector space.

3 Applications and Examples

We use this formulation to study the effect of the Markov time \( t \) as a resolution parameter in multiscale community detection under the Markov Stability framework.

A constructive network: non-hierarchical structure (Fig. 1).

Fig. 1: (a) The graph with \( n = 12 \) nodes; (b) Visualization of the vectors using the first two components at Markov time \( t = 2 \) and \( t = 5 \). For this network, the optimum of the maximum length vector partitioning problem changes from a 3-way partition (blue) to a 2-way partition (red) as time gets larger and the vectors \( x_i \) become closer. Note that due to symmetry, only six (out of the twelve) vectors \( x_i \) are visible on this projection for this graph.

A network from real data: when spectral clustering fails. A \( k \)-nearest-neighbour graph \( (k = 10) \) with \( n = 13782 \) nodes is constructed from the images of digits 4 and 9 from the MNIST dataset [3] (Fig. 2).

The 5th International Workshop on Complex Networks & Their Applications. Nov. 30 - Dec. 02, 2016, Milan (Italy)
Fig. 2: Visualization of the network of MNIST 4-9 digits using the first two components of the $x_i$ vectors and the partitions given by: (a) the ground truth; (b) vector partitioning at Markov time $t = 1$; and (c) vector partitioning at Markov time $t = 100$. The result of vector partitioning at $t = 100$ is given by the sign of the second eigenvector and coincides with the Fiedler bipartition from spectral clustering. The partition at $t = 1$ gives community assignment much closer than those at $t = 100$ (or spectral clustering).

4 Discussion

Our formulation differs from the work in [4] where the decomposition is based on the eigenvalues of the modularity matrix, whereas we obtain our vectors via a generalized eigenvalue problem and all the eigenvalues are positive. Furthermore, our vector partitioning problem is parametrically dependent on Markov time, i.e., the components of the vectors shrink according to Markov time yet the different rates are determined by their corresponding eigenvalues. This feature allows for the exploration of changes in optimal groupings at different resolutions leading to potential multi-scale community detection.

Acknowledgements

This work is funded by the European Union 7th Framework Programme for research, technological development and demonstration under grant agreement no. 607466. MB acknowledges support from the EPSRC through grant EP/N014529/1.

References

The ground truth about metadata and community detection in networks

Leto Peel\textsuperscript{1,2}, Daniel B. Larremore\textsuperscript{3}, and Aaron Clauset\textsuperscript{4,5,3}

\textsuperscript{1} ICTEAM, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
\textsuperscript{2} naXys, Université de Namur, Namur, Belgium
\textsuperscript{3} Santa Fe Institute, Santa Fe, NM 87501, USA
\textsuperscript{4} Department of Computer Science, University of Colorado, Boulder, CO 80309, USA
\textsuperscript{5} BioFrontiers Institute, University of Colorado, Boulder, CO 80309, USA

1 Summary

Community detection is a fundamental task of network science that seeks to describe the large-scale structure of a network by dividing the network’s vertices into communities, blocks, or groups, based only on the pattern of edges. It is common to evaluate the performance of community detection algorithms by their ability to find so-called ground truth communities. This works well in synthetic networks with planted communities because such networks’ links are formed explicitly based on the planted communities. However, there are no planted communities in real world networks. Instead, it is standard practice to treat some observed discrete-valued node attributes, or metadata, as ground truth. Here, we show that metadata are not the same as ground truth, and that treating them as such induces severe theoretical and practical problems. We prove that no algorithm can uniquely solve community detection, and we prove a general No Free Lunch theorem for community detection, which implies that no algorithm can perform better than any other across all inputs. However, node metadata still have value and a careful exploration of their relationship with network structure can yield insights of genuine worth. We illustrate this point by introducing two statistical techniques that can quantify the relationship between metadata and community structure for a broad class models. We demonstrate these techniques using both synthetic and real-world networks, and for multiple types of metadata and community structure.

2 Metadata labels are not ground truth communities

The merit of using metadata labels as ground truth communities is that if there is correspondence between communities and metadata, it tells us that there is a relationship between the network structure and the metadata, while at the same time implying that the community detection algorithm is identifying useful communities. However, networks can have many plausibly “good” partitions \cite{11}, so when communities and metadata do not match it is not necessarily because the community detection algorithm does not perform well. In fact, when analyzing real-world networks, there are four possible reasons for mismatch between metadata and communities: (i) metadata do not relate to network structure, (ii) communities and metadata capture different aspects of network structure,
as shown in Fig 1 (iii) the network contains no real structure, or (iv) the algorithm performed poorly. Despite these possibilities, typically the assumption is that (iv) is the only possible cause.

Furthermore, treatment of metadata labels as ground truth communities, supposes that a single set of true communities really exists. However, there are many competing definitions of community, each with its merits, and this multiplicity is not without consequences. We prove two theorems which show that a unique solution for community detection does not exist and that no community detection algorithm can perform better than any other on average.

3 Methods to explore metadata-community relationships

While metadata labels should not serve as ground truth to calibrate community detection algorithms, the relationship between network structure and metadata can be meaningful. In many cases, metadata may enable us to learn about the processes responsible for link formation or dissolution. To better diagnose the actual relationship between metadata and structure, we briefly describe two methods below. We demonstrate these methods with respect to the popular stochastic blockmodel [2] [4], but both can be adapted to the broad class of generative network models.
First, we introduce the blockmodel entropy significance test to determine whether or not the metadata and network structure are related. In this test we calculate the entropy of the network, given the metadata, by fitting a stochastic blockmodel using the partition given by the metadata labels. We compare this entropy against a null distribution of entropies calculated by randomized metadata labels. Using this test we can determine the probability that the observed metadata labels were assigned at random relative to the network, addressing case (i) above. We identify relationships between metadata and network structure in real data and demonstrate cases where there exists a significant relationship between multiple sets of metadata and the network structure.

Second, we introduce a new stochastic blockmodel, the neoSBM, which explores the transition through partitions between metadata and best-fit communities. This method addresses case (ii) above. Specifically, the neoSBM chooses whether or not each node is assigned to its metadata community or if it is free to choose its own community at a cost. By varying this cost, the neoSBM effectively explores the space of partitions to find a path between the metadata and community partitions. The type of path tells us about how the two partitions are related. A smooth path between the two indicates that the metadata is close to the global optimum and suggests that they represent the same aspect of the network structure. On the other hand, the presence of a sharp phase transition in the path suggests that the metadata is at (or close to) a different local optimum, which we can interpret as a different aspect of the network structure. Figure 1 shows the partition space and likelihood under the SBM for the karate club network [5]. Here we see that the metadata corresponds to a local optima representing the assortative group structure of the network, while the global optimum captures a core-periphery structure. Both are relevant and interesting aspects of the network.

References

Finding Multiple Core/Periphery Structure with Random Walks

Sadamori Kojaku\textsuperscript{a} and Naoki Masuda\textsuperscript{b}

Department of Engineering Mathematics, University of Bristol, Bristol, U. K.
\textsuperscript{a}sadamori.koujaku@bristol.ac.uk
\textsuperscript{b}naoki.masuda@bristol.ac.uk

Abstract. We introduce a novel algorithm to find multiple core/periphery structure in networks. Previous methods to detect core/periphery structure have focused on the case in which a network consists of a single group of core nodes and the corresponding peripheral nodes. However, networks may have multiple pairs of core and periphery. To find the multiple core/periphery pairs, we extend modularity, a quantity function for community detection in networks, by incorporating the contention that random walkers on peripheral nodes are likely to move to core nodes. Our algorithm can automatically determine the number of core/periphery pairs and also is scalable to large networks. We tested our algorithm on several synthetic and real networks.

1 Introduction

Core/periphery structure is a mesoscale characterisation of networks in which a core is a set of densely interconnected nodes, and a periphery is a set of nodes that mainly connect to the core. Recently, core/periphery structure is found in various networks such as social networks, ecological networks and biological networks \cite{4,6,8}.

The core/periphery structure has been studied under the assumption that a network consists of a single core and the corresponding periphery \cite{3,5,7,8,9}. However, a network may have multiple pairs of core and periphery. Here we propose a scalable algorithm for detecting multiple cores and the corresponding peripheries in a given network.

2 Methods

Suppose that we start a random walk from a node selected with the equal probability in a graph. Since peripheral nodes connect mainly to the corresponding core nodes, the walkers initially located in the periphery are likely to go to the core after one move. The walkers in the core are less likely to go to a peripheral node because of dense interconnections between the core nodes. Therefore, the density of walkers after one move is expected to increase in the core and decrease in the periphery. On the basis
of this property, we translate changes in the density of walkers after one move into a quality similar to modularity. We use Louvain algorithm [2], a widely used algorithm for maximising modularity, to maximise the proposed quantity. The maximisation yields groups of core nodes and one group of non-core nodes. Then, we assign each non-core node to one group of core nodes as its periphery such that the nodes in the selected group have the largest summed number of edges with the non-core node.

3 Results

We tested the proposed algorithm on several artificial and empirical networks. Here we only report the results for an empirical network. Figure 1 shows the network of political blogs about the United States president election in 2004, in which a node is a blog with one of the two political leanings (liberal or conservative) [1]. Two blogs are connected if one cites the other blog. We regard that the network is undirected. An algorithm to detect single core/periphery [3] found one core that mainly included nodes in the liberal and some nodes in the conservative. The original modularity maximised by Louvain algorithm [2] and the proposed algorithm found two large groups that reasonably matched the two political leanings. The proposed algorithm also found the densely connected core nodes of each group.

4 Conclusions

We introduced an extension of modularity to find multiple core/periphery structure in networks based on the tendency that random walkers are likely to move to core nodes after one move. For the network of political blogs, the proposed algorithm has found two pairs of core and periphery, each of which was formed by liberal and conservative blog posts.

Future work includes (a) applying a similar extension to other community detection methods to construct algorithms to find core/periphery structure, (b) investigation of multiple core/periphery organisation in other networks, and (c) analysing temporal changes in core/periphery structure.

References


Fig. 1: Network of political blogs on the United States presidential election in 2004. The colours of the nodes show the political leaning of the associated blogs in (a), and the membership of the detected groups in (b), (c) and (d). The filled and empty circles in (c) and (d) represent detected core and peripheral nodes, respectively.
Patient pathways reveal network flows with memory

Gergely Palla¹, Nora Pall², Anna Horvath³, Katalin Molnar³, Tamas Kovats⁴, Gyorgy Surjan⁴, Tamas Vicsek⁵, and Peter Pollner¹

¹ MTA-ELTE Statistical and Biological Physics Research Group, Budapest, Hungary, pollner@angel.elte.hu
² Regional Science Center, Faculty of Science, Eötvös Loránd University
³ 3rd Department of Internal Medicine, Semmelweis University
⁴ National Healthcare Service Center
⁵ Department of Biological Physics, Eötvös Loránd University Budapest, Hungary

1 Introduction

Spreading and diffusion processes on networks are often modelled by random walks, where the current position of the walker fully determines the transition probabilities for the next step. However, according to recent results the flows on e.g., transportation networks, citation networks, etc., have memory and cannot be treated as first order Markov-processes [3, 4]. This can have serious consequences on the ranking between the nodes and may lead to a community structure that is hard to extract with traditional methods [3, 2].

2 Patient flow network

In this work we show that in a somewhat similar fashion, the flow of medical patients between different institutions of the health care system can be considered as a network with memory as well. Our study is based on a detailed database containing the records of hospital visits in a 13 year period of 11 million anonymised patients. Since a considerable fraction of the patients had several hospital visits in the studied time interval, the re-appearance of a previously recorded patient in the same- or in a different hospital can be treated as a flow between the institutions. Beside the dates of the hospital visits the ICD-9 classification of the diagnosed diseases were also recorded, thus, the pathways characteristic to a given disease could be also studied.

According to our results the patient flow network between health care institutions has memory both on the level of individual diseases and also when patients with different ICD-9 classification codes are aggregated. Therefore, the process shaping the patient pathways cannot be modelled by a traditional random walk on a network between the hospitals, and is more analogous to higher order Markov processes.

3 Instability of pathes of patients with an autoimmune disease

Beside the general aspects related to the memory of the processes, we present results on a rarely accessible application field. We study the pathways of a patient group with
an autoimmune disease (relapsing polychondritis). This disease has a marked variability in the symptoms. The complex processes that are induced by the disease lead to a wide range of clinical manifestations and nontrivial patient pathways on the network of different medical centres. The first studies pointing in the direction of the medical applications of the involved datasets have been already carried out [1]. Now we present the analysis of the pathway structure (see Fig. 1): we measure the high instability of the paths and the separability of patient histories according to the stages of the disease.

Fig. 1: Pathways of patients with relapsing polychondritis. Nodes represent clinical centres of different professions. Each column indicates a time step along a pathway and centres with the same profession are aligned into the same rows. Links between the nodes represent patients that have visited the connected medical centres at consecutive time steps.

References

Sample Space Reducing processes (SSRP) offer an alternative new mechanism to understand the emergence of scaling in countless processes [1]. SSRPs are processes that reduce their sample space as they progress over time. In their simplest form they can be depicted by the following process. Imagine a staircase like the one shown in figure 1(a, b). Each state $i$ of the system corresponds to one particular stair. A ball is initially ($t=0$) placed at the topmost stair $N$, and can jump randomly to any of the $N-1$ lower stairs in the next timestep with a probability $1/(N-1)$. Assume that at time $t=1$ the ball landed at stair $i$. Since it can only jump to stairs $i'$ that are below $i$, the probability to jump to stair $i' < i$ is $1/(i-1)$. The process continues until eventually stair 1 is reached; it then halts. The corresponding power law exponents have been shown to be related to noise –probability to jump upwards– levels in the process. Moreover, the emergence of scaling is not limited to the simplest SSRPs, but holds for a huge domain of stochastic processes that are characterised by non-uniform prior distributions [2]—see fig 1c. Zipf’s law turns out to be an attractor distribution that holds for practically any

Fig. 1: SSR processes: a standard sampling process (a) reproducing the uniform distribution sampled. (b) a standard SSR process: as soon as state $i$ is hit, only states $k < i$ are reachable, the obtained histogram of visits is Zipf’s law, i.e., a power-law with an exponent $= -1$. (c) SSR process sampling a non-uniform prior probability distribution still leads to Zipf’s law in many cases, regardless the functional shape of the underlying distribution.
SSRP, irrespective of the details of the stochastic system at hand, i.e. irrespective of their prior distributions. As a consequence, processes of diffusion on networks where (a) target(s) have been defined, can be understood [2]. Here Zipf’s distributions of node visiting frequencies appear generically, regardless of the weight- or degree distribution of the network. We call diffusion processes on DAG structures targeted diffusion, since, in this type network, diffusion is targeted towards a set of target or sink nodes, see figure 2e. Specifically, the SSRP framework enables to explain the associated scaling laws in node visit distributions. Remarkably, the emergence of scaling in the traffic properties is largely independent of the topology and weight distribution of the edges, as predicted by the theory of SSRPs –see fig 1(a–e). The presence of cycles can be properly interpreted as playing the same role as noise in SSRPs and, accordingly, determine the scaling exponents –see fig 2(a,f). These results could be relevant for a series of applications in traffic-, transport- and supply chain management.

References

Poker Games on Complex Networks

Marco Alberto Javarone

Dept. of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy
marcojavarone@gmail.com

1 Introduction

Poker represents one of the major challenges for artificial intelligence and mathematics and, in addition, it is a topic of interest also for psychologists, economists and sociologists due to its wide diffusion. One of its most controversial aspects, mainly caused by the utilization of money, is related to its nature, i.e. ‘skill game’ or gambling. The related answer has not yet been solved, although it has a wide list of implications. Moreover, all efforts made to define algorithms and strategies in the context of artificial intelligence are obviously based on the confident belief that computing skills are crucial for succeeding in this game. There are two main formats for playing poker: tournament and ‘cash game’. The former entails players pay an entry fee that goes into the prize pool plus a fee to play, receiving an amount of chips. Then, top players share the prize pool. Instead, playing poker in the ‘cash game’ format entails to use real money during the challenge. So, in this last case, players can play until they have money and, although there are no entry fees to pay, a fraction of each pot is taxed. In the work [1], the author defined a model for representing poker challenges, focusing on tournaments, in order to study the role of rationality. As result, he found that the nature of poker (i.e. gambling or skill game) does not depend on its rules but on the players’s behavior, then identifying rationality as a key ingredient to succeed. Later on, in [2], a mean-field approximation of the previous model (i.e. [1]) has been analyzed. In this second work, it has been shown that when agents change strategy over time, by imitating their opponents, there is a neat difference between tournaments and ‘cash game’ challenges. The former is confirmed to be a skill game, while the latter seems to be closer to gambling. Here, we aim to analyze a model for representing the dynamics of ‘cash game’ challenges in complex networks. In particular, the interest towards this investigation lies in a phenomenon that we can observe in online Poker platforms: some players face multiple opponents at time. This strategy is defined ‘multi-tabling’, i.e. one user takes part to a very high number of challenges, in different virtual tables, at the same time. Remarkably, there are players able to manage more than 20 tables at time; so, as we can imagine, this strategy can become really profitable only in presence of high quality Poker skills. Therefore, complex networks might constitute the right framework for studying these dynamics. In particular, we analyze if the network topology affects the gain of rational players, when Poker challenges are based on the ‘cash game’ structure.
2 Model

In the proposed model, we consider a population composed of Rational (i.e. $R$) and Irrational (i.e. $I$) agents playing 'cash game' challenges. As for the previous models (i.e. [1] and [2]) numerical simulations are based on 'heads-up' challenges, i.e. they involve two agents per table. Rational agents have a very high probability to succeed when they receive high value cards but, at the same time, they perform a high number of 'fold' actions. The latter entails that a player prefers to lose a small amount of money than risking a higher fraction of her/his bankroll (i.e. capital). We performed numerical simulations in three different cases: agents arranged on regular lattices (with continuous boundary conditions), agents randomly arranged in scale-free networks, and agents arranged in scale-free networks placing rational agents in the hubs (i.e. nodes with high number of connections). The dynamics of the model are very simple: at each time step one randomly chosen agent plays a Poker challenge with its neighbors. Agents are memory-aware (e.g. see [3]), i.e. they save their gain over time. In doing so, after a number of time steps we can evaluate if, on average, Rational agents are richer than Irrational ones, or vice versa.

3 Results and Discussion

Here, we present the preliminary results of the proposed model, achieved on varying the density of rational agents in the population. Figure 1 reports results of numerical simulations for all three cases: regular lattice, scale-free with random distribution of agents, and scale-free with rational agents located in the hubs of the network.

This result clearly indicates that Rational agents can take profit only when arranged in scale-free networks and, in particular, when located in the hubs. In addition, as expected, increasing the number of rational agents, their gain decreases. Notably, since two agents of the same kind (i.e. both Rational or both Irrational) have the same probability to succeed when face each other, in the limit $\rho_r \to 1$ the average gain of Rational agents goes to zero. Therefore, we can state that a multi-tabling strategy can be profitable only when there are no too many rational agents in the population.

As for future work, we aim to consider other complex topologies to corroborate the results achieved in this preliminary work.

Acknowledgments

This work was supported by National Group of Mathematical Physics (GNFM-INdAM).

References

Fig. 1: Average gain of rational agents (i.e. $G_r$) in function of their density $\rho_r$ in the population. Red line represents results achieved by arranging agents in a regular lattice (with continuous boundary conditions). Blue line represents results achieved by arranging agents in scale-free networks, without to consider if rational or irrational agents are located in hubs. Green line represents results achieved in scale-free networks, placing rational agents in the hubs of the network.
Epidemic threshold in hierarchical modular network models

Ali Safari and Paolo Moretti
Institute of Materials Simulation (WW8), Dr.-Mack-Straße 77, 90762 Fürth, Germany,
ali.s.safari@fau.de
paolo.moretti@fau.de

Abstract. Hierarchical modular networks are frequently encountered in biological systems and exhibit singular spectral properties, such as absence of a well-defined spectral gap and pronounced eigenvector localization. We study the relationship between such unusual spectral features and the nature of the epidemic threshold in hierarchical modular networks. By means of numerical simulation and theoretical arguments, we show that the standard relationship between the threshold value of the spreading rate in the susceptible-infected-susceptible dynamics and the largest eigenvalue of the adjacency matrix does not hold in this case, and we discuss a possible generalization of it, which accounts for the unique spectral features of hierarchical modular connectivity patterns.

1 Background

The mechanism leading to the onset of epidemic spreading in complex networks has been the subject of extensive investigation over the past years (See [1] for a review). Particular attention has been recently devoted to networks exhibiting a given degree of eigenvector localization. It was initially proposed [2, 3] that in localized networks the epidemic threshold should be replaced by an interval of the spreading rate \( \lambda \), in which short-lived active states replace an active steady state (a Griffiths phase). While this view has been challenged in the case of scale-free networks [4], a similar phenomenology [5] is recovered in hierarchical modular networks (HMNs, see Fig. 1). HMNs are relevant in the modeling of biological networks such as the human connectome. In HMNs, pairs of modules are connected establishing a fixed number of links \( \alpha \) between their nodes, thus forming super-modules that are in turn connected pairwise with the same \( \alpha \). This process is repeated across several hierarchical levels, until the whole network is connected. HMNs built according to this patterns are thus parametrized by the connectivity strength \( \alpha \) defined above. They are characterized by a finite topological dimension and exhibit remarkable spectral properties: (i) Lack of properly defined spectral gap. While a separation exists between the two largest eigenvalues of the adjacency matrix \( A \) (as guaranteed by the Perron-Frobenius theorem), it is not large enough to justify analytic expansions that take into account only the principal eigenvalue \( \Lambda_1 \) and its corresponding eigenvector; (ii) Localization, measured by the inverse participation ratio (see for instance [2]), which extends to all the eigenvectors in the higher spectral edge of \( A \).
Fig. 1: Schematic representation of a hierarchical modular network (HMN). Blue boxes represent densely connected modules. Modules are grouped pairwise, forming super-modules, which are themselves grouped in pairs in an iterative hierarchical fashion. At every hierarchical level, modules are paired by establishing a fixed number of links ($\alpha = 2$ in figure). Each link connects two randomly chosen nodes from each module/super-module.

2 Results

We investigate the emergence of a non-zero epidemic-like dynamic threshold $\lambda_c$ in HMNs for values of $\lambda$ larger than the inverse of the largest eigenvalue $\Lambda_1$ of $A$, as first observed in [5]. Using theoretical arguments and extensive numerical simulations of susceptible-infected-susceptible dynamics, we are able to relate this phenomenology to the localization properties of HMNs and provide a generalized predictive criterion, which reduces to the well-known $\lambda_c = 1/\Lambda_1$ rule in the standard case. While the inverse of largest eigenvalue $\Lambda_1$ provides a lower bound for the epidemic threshold in any network, it was shown in [5] that such lower bound deviates significantly from the actual $\lambda_c$ in HMNs: in other words the lower bound alone does not provide an accurate prediction of $\lambda_c$. By means of analytical calculations and extensive numerical simulations, here we show that in HMNs eigenvector localization makes it impossible for activity to be sustained at $\lambda_c = 1/\Lambda_1$, and an expansion including other eigenvalues in the spectral edge is required. We demonstrate that these eigenvalues scale with the structural parameter $\alpha$, finally obtaining a scaling law for the epidemic threshold in the form $\lambda_c \sim 1/\alpha$.

References

Emergence of Resilience in Growing Networks by Range-limited Interedications

Yukio Hayashi

Japan Advanced Institute of Science and Technology, Ishikawa, Japan
yhayashi@jaist.ac.jp

Several years ago, by percolation analyses, it has been discovered that onion-like topological structure with positive degree-degree correlations gives the optimal robustness of connectivity even for the conventional most serious case against the intentional hub attacks on scale-free (SF) networks [8][9] to which many real social, technological, and biological networks belong. In the state-of-the-art, for generating such optimal robust onion-like networks, there is only a method by not whole rewiring [10] but incrementally growing [3][4]. On the other hand, new threats of network attacks with severer damage than the hub attacks appear recently [5, 6]. They are Collective Influence (CI) attack [5] considered for a global optimization to identifying the most influence nodes in information spreading and Belief Propagation (BP) attack [6] derived from a message passing approximation algorithm rooted by the spin glass model in statistical physics for the combinatorial Feedback Vertex Set problem. These attacks based on destruction of loops and a Zeta function of graphs [2] give theoretical motivations to associate long loops with network connectivity. Thus, we inversely take into account the new threats and propose another self-organized method for incrementally growing resilient onion-like structure by enhancing loops as follows.

In prohibiting multiple links between two nodes, we introduce pairs of random and long distance attachments (RLD-A) with even number \( m \) links from an added new node at each time step. The pairs of attachments form interwoven loops at the new node. Furthermore for distributed information systems, we consider the range-limited approximation of RLD-A as random and intermediated attachments (MED): Instead of the furthest node in RLD-A, we select a distant node to the extent of a few hops via intermediations from the randomly chosen pair node. Intermediations of \( \mu \) hops mean attachments to the \( \mu + 1 \)-th neighbors of the randomly chosen pair node. For the proposed growing networks, we compute the robustness index \( R \equiv \frac{1}{\pi} \sum_{q=1}^{1/N} S(q) \) and assortativity \( r \) as the Pearson correlation coefficient for degrees [7], where \( S(q) \) denotes the number of nodes included in the giant component (GC) after removing \( qN \) nodes by High Degree Adaptive (HDA), CI for \( l = 3 \) layer [5], and BP attacks with inverse temperature \( x = 7 \) and 100 rounds of message-passing [6]. As shown in Figs. 1ac, our networks by RLD-A and MED\( \mu \) for reasonable \( \mu = 3, 4 \) and \( m = 4 \) have both high \( R > 0.3 \) and \( r > 0.2 \), thus they become onion-like networks even without any operations to enhance degree-degree correlations. Figure 1b shows the sudden breakdowns of GC by BP attacks (bluish lines). The robustness of SF networks by BA model [11] is weaker with smaller \( R \approx 0.2 \) (but not shown in Fig. 1b to avoid crowded lines). Moreover, our networks have efficient \( O(\log N) \) path length as shown in Fig. 1d. As a virtual test for exploring future design of infrastructure networks, we also investigate the resilience in the growing from real networks by the intermediations, and obtain the drastic improvement of robustness.
Fig. 1: Basic properties for the proposed growing networks from an initial complete graph of $N_0 = m$ nodes. a) Almost constant strong robustness. b) Comparison between the proposed and BA models at $N = 5000$ for the breaking of GC against HDA, CI3, and BP attacks. c) Degree-degree correlation measured by assortativity $r$ for growing size $N$. d) Small-world-ness as $O(\log N)$ of the averaged path lengths over the paths of the minimum hops between two nodes.

References

Finding the Most Influential Nodes in Pinning Controllability of Complex Networks

Ali Moradi Amani, Mahdi Jalili, Xinghuo Yu, and Lewi Stone

Abstract. Identifying the best drivers (i.e. the nodes to apply the control signals in a large complex network), which gives the fastest synchronization to the reference state is a challenge in pinning control of a network. In this manuscript, we introduce a novel method that gives first-order approximation for the importance of nodes in pinning control. A spectral measure (the largest eigenvalue of the augmented Laplacian matrix divided by the smallest one) is considered as pinning controllability metric. We develop this method based on the sensitivity analysis of the Laplacian eigenratio, resulting to scoring the nodes based on their importance in pinning control. The method is rather simple to compute and need a single eigen-decomposition of the Laplacian matrix of the connection graph.

Introduction

A challenge in the context of pinning controllability of complex networks is to find the best driver nodes, which give the best pinning control performance, e.g. fastest synchronization to the reference state. If one has no prior knowledge on the importance of nodes in pinning control, the easiest approach is to choose the most central nodes (e.g., those with the highest degree, betweenness or closeness centrality) as drivers. These methods, although being efficient, often results in not efficient pinning controllability, and one can further optimize the performance [4, 5, 6].

In this manuscript, we perform a sensitivity analysis on the graph Laplacian and introduce an index to rank the nodes based on their influence on the pinning controllability. The introduced measure, denoted as Eigenratio Sensitivity Index (ESI), is based on the eigenvectors corresponding to the largest and smallest eigenvalues of the Laplacian matrix, and one only needs a single eigen-decomposition to compute it.

Eigenratio Sensitivity Analysis

Definition 1. [2]. For each node $i$ of the undirected network $(V, E)$, the Eigenratio Sensitivity Index (ESI) is defined as $ESI(i) = (x_i^N)^2$, where $x_i^N$ represents the $i$th element of $x^N$, the eigenvector corresponding to the largest eigenvalue ($\lambda_N$) of the Laplacian matrix of the graph.

Theorem 1. [2]. In the undirected network $(V, E)$, the node with maximum ESI has the strongest influence on the eigenratio of the Laplacian matrix.

Simulation Results

In this section we apply ESI on a number of synthetic networks and compare its performance with a number of heuristics methods. As network modes, we consider scale-free...
Fig. 1: Accuracy of ESI (solid), maximum degree (dash-dot), maximum betweenness centrality (dot) and maximum closeness centrality (dashed) in finding the most influential driver node in scale-free network with $N = 1000$ nodes with A) $B = 0$, B) $B = 5$, and C) $B = 10$ (as $B$ increases, the heterogeneity of the network decreases). Average degree of the networks varies from 2 to 20 and the results show the accuracies over 100 realizations.

Fig. 2: Accuracy of ESI (solid), maximum degree (dash-dot), maximum betweenness centrality (dot) and maximum closeness centrality (dashed) in finding the top-$T$ most influential driver nodes in networks with $N=1000$ nodes. Networks have scale-free structure with $B = 0$ and A) $m = 5$, B) $m = 10$ and C) $m = 20$ as well as $B = 5$ and D) $m = 5$, E) $m = 10$ and F) $m = 20$ ($m$ is the average degree. In addition, as $B$ increases, the heterogeneity of the network decreases). Results are averaged over 100 realizations.
proposed in [3] and the model proposed by Watts and Strogatz in their seminal work [7].

The ESI is also applied in finding the most influential node in cooperative frequency control of distributed generation system [1]. A power network including 200 generation nodes (including synchronous generators and renewable resources) is considered. Considering operational constraints, it is assumed that only 50 of these generation sources are candidates to be frequency leader of the network, while others will be in droop mode, following the leader. The proposed technique is used to select the leader. Figure 3 compares the accuracy of ESI in finding the best leader with other heuristic measures. To calculate the accuracy, the nodes are selected as leader one-by-one and are sorted based on their real effect on the frequency dynamics. Top-k nodes are compared with their counterparts achieved using ESI and other heuristics.

![Accuracy in finding top-k nodes](image)

Fig. 3: Accuracy in finding top-k nodes with maximum influence on the stability of distributed generation system when the communication data has a scale-free structure. Results are averaged over 100 realizations.

**References**