Springer Proceedings in Complexity

Fakhteh Ghanbarnejad
Rishiraj Saha Roy
Fariba Karimi
Jean-Charles Delvenne
Bivas Mitra *Editors*

Dynamics On and Of Complex Networks III

Machine Learning and Statistical Physics Approaches

Springer

Springer Proceedings in Complexity

Springer Proceedings in Complexity publishes proceedings from scholarly meetings on all topics relating to the interdisciplinary studies of complex systems science. Springer welcomes book ideas from authors. The series is indexed in Scopus. Proposals must include the following: – name, place and date of the scientific meeting – a link to the committees (local organization, international advisors etc.) – scientific description of the meeting – list of invited/plenary speakers – an estimate of the planned proceedings book parameters (number of pages/articles, requested number of bulk copies, submission deadline). submit your proposals to: christopher.coughlin@springer.com

More information about this series at http://www.springer.com/series/11637

Fakhteh Ghanbarnejad • Rishiraj Saha Roy
Fariba Karimi • Jean-Charles Delvenne
Bivas Mitra
Editors

Dynamics On and Of Complex Networks III

Machine Learning and Statistical Physics Approaches

Springer

Editors
Fakhteh Ghanbarnejad
Institute of Theoretical Physics
Technical University of Berlin
Berlin, Germany

Rishiraj Saha Roy
Max Planck Institute for Informatics
Saarbrücken, Germany

Fariba Karimi
Department of Computational Social Science
GESIS
Leibniz Institute for the Social Science
Köln, Germany

Jean-Charles Delvenne
Université Catholique de Louvain
Louvain-la-Neuve, Belgium

Bivas Mitra
Department of Computer Science and Engineering
Indian Institute of Technology Kharagpur
Kharagpur, India

ISSN 2213-8684 ISSN 2213-8692 (electronic)
Springer Proceedings in Complexity
ISBN 978-3-030-14682-5 ISBN 978-3-030-14683-2 (eBook)
https://doi.org/10.1007/978-3-030-14683-2

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Recently “network science” has been bridging various disciplines like mathematics, physics, biology, chemistry, computer science, ecology, and the social sciences. This is mainly due to its wide perspective in modeling the structure and dynamics of complex systems, both natural and man-made, with different, large, or even multiple scales. Some examples include genetic networks, food web, trade networks, the World Wide Web (WWW), collaboration networks, power grids, and air traffic networks.

The primary aim of the workshop series on *Dynamics on and of Complex Networks* (DOOCN) is to systematically explore the statistical dynamics “on” and “of” complex networks that prevail across a large number of scientific disciplines. *Dynamics on networks* refers to dynamical processes which evolve on networks and their evolution, which is impacted by their underlying topology. On the other hand, *dynamics of networks* refers to the changes occurring in the topology due to various interactions. The first DOOCN workshop (DOOCN-I) took place in Dresden, Germany, 2007, as a satellite workshop of *the European Conference on Complex Systems* (ECCS). After the success of DOOCN-I, new editions were organized as satellites of *E*CCS/CCS in Jerusalem (2008), Warwick (2009), Lisbon (2010), Vienna (2011), Barcelona (2013), Lucca (2014), Amsterdam (2016), and Thessaloniki (2018) and also as satellites of *the International School and Conference on Network Science* (NetSci) in Zaragoza (2015) and Indianapolis (2017). Details of the workshop series are available at doocn.org.

Eminent speakers of the workshops in recent years included (alphabetically ordered) Katharina Zweig (TU Kaiserslautern, Germany), Renaud Lambiotte (University of Oxford, UK), Yamir Moreno (University of Zaragoza, Spain), Sayan Pathak (Microsoft Research Redmond, USA), Ginestra Bianconi (Queen Mary University of London, UK), Constantine Dovrolis (Georgia Institute of Technology, USA), Martin Rosvall (Umea University, Sweden), Frank Schweitzer (ETH Zurich, Switzerland), Krishna Gummadi (Max Planck Institute for Software Systems, Germany), Ciro Cattuto (ISI Foundation, Italy), Markus Strohmaier (RWTH Aachen University, Germany), and Matthieu Latapy (UPMC, Paris).

Notably, the workshop organizing committee has published two book volumes from the selected talks of the series (2009: https://goo.gl/tmqPYm) and (2013: https://goo.gl/GQkfEp). The first volume aimed to show how complex network theory is being successfully used by researchers to tackle numerous difficult problems in various domains and included three parts addressing applications of complex networks in biological, social, and information sciences. The second volume aimed to put forward burgeoning multidisciplinary research contributions that combine methods from computer science, statistical physics, econometrics, and social network theory toward modeling time-varying social, biological, and information systems. This volume included three parts: (1) online social media, the Internet, and the WWW, (2) community analysis, and (3) diffusion, spreading, mobility, and transport.

The third book volume is the present one, edited by the DOOCN 2017 organizers, and aims to focus on this specific topic: "Machine Learning and Statistical Physics". Recently, machine learning (ML) techniques have been used to model dynamics of massive complex networks generated from big data and various functionalities resulting from the networks. It has become clear from the past DOOCN workshop editions that modeling large-scale dynamic networks, such as mobile adhoc networks and societal opinion networks, has gained enormous relevance in the landscape today. The advent of big data technologies, which allow effective acquisition and processing of massive amounts of unstructured data, further promises to improve the effectiveness and cross-fertilization of ML and network science. This motivated us to focus on this area of significant recent interest in our last workshop editions. A key feature of the DOOCN workshops is that each year, an exciting theme is chosen; some recent themes include "Big Data" (2014), "Computational Aspects of Big Data" (2015), "Mining and learning for complex networks" (2016), "Machine learning and statistical physics" (2017), and "Machine learning for complex networks" (2018). This volume presents a mix of very relevant reviews of important works in the field and gives the reader an up-to-date picture of the state of the art. This edition also contains independent research reports.

This book volume consists of three major parts. The contributions in the first part focus on network structure, with three chapters. In the first chapter "An Empirical Study of the Effect of Noise Models on Centrality Metrics", Sarkar et al. conducted an empirical study of how different noise models affect the network structure, precisely the ranking of centrality metrics. The analysis presented in this chapter reveals that the stability of the ranking varies according to the structure of the network, the noise model used, and the centrality metric to be computed. In the second chapter "Emergence and Evolution of Hierarchical Structure in Complex Systems", Siyari et al. investigated the following key questions in the context of modeling the emergence and evolution of hierarchical structure in complex systems: (a) How do key properties of emergent hierarchies, like depth of the network, centrality of each module, and complexity of intermediate modules, depend on the evolutionary process that generates the new targets of the system? (b) Under what conditions do the emergent hierarchies exhibit the so-called hourglass effect? (c) Do intermediate modules persist during the evolution of hierarchies? In the third chapter

"Evaluation of Cascading Infrastructure Failures and Optimal Recovery from a Network Science Perspective", Warner et al. reviewed the network science literature in order to create a hypothesis for the recovery of infrastructure systems. They represented the cascade of infrastructure systems through networks and simulated perturbations within singular and discussed how these impact multiple layers of an interconnected network.

Part II of the book volume focuses on network dynamics and it spans over four chapters. In the fourth chapter "Automatic Discovery of Families of Network Generative Processes", Menezes and Roth first reviewed the principles, efforts, and emerging literature in this direction, which is aligned with the idea of creating artificial scientists. Next, the authors developed an approach to demonstrate the existence of families of networks that may be described by similar generative processes. In the fifth chapter "Modeling User Dynamics in Collaboration Websites", Kasper et al. presented several approaches to deepen the understanding of user dynamics in collaborative websites. Inevitably, these approaches are quite heterogeneous and range from simple time-series analysis toward the application of dynamical systems and generative probabilistic methods. In the sixth chapter "Interaction Prediction Problems in Link Streams", Arnoux et al. addressed the problem of predicting future interactions, which is traditionally addressed by merging interactions into a graph or a series of graphs, called snapshots. However, in this chapter, authors formalized interactions within the link stream framework, which makes it possible to fully capture both temporal and structural properties of the data. In the seventh chapter "The Network Source Location Problem in the Context of Foodborne Disease Outbreaks", Horn and Friedrich introduced the source identification problem in the context of foodborne disease outbreaks based on basic practical knowledge of food supply networks and the foodborne disease contamination process.

The third part of the book volume focuses on theoretical models and applications with three chapters. In the eighth chapter "Network Representation Learning using Local Sharing and Distributed Graph Factorization (LSDGF)", Pandey proposed a distributed algorithm for network representation learning (NRL), which learns matrix factorization of a given network in which a node utilizes only the information available at its neighboring nodes and connected nodes for exchanging feature vectors dynamically. The performance of the proposed algorithm is evaluated by the learning of first-order proximity, spectral distance, and link prediction. In the ninth chapter "The Anatomy of Reddit: An Overview of Academic Research", Medvedev et al. explored one of the most popular social media platforms, Reddit. They developed a suite of methodologies to extract information from the structure and dynamics of the Reddit system. In the tenth chapter, "Learning Information Dynamics in Online Social Media: A Temporal Point Process Perspective", Samanta et al. proposed two models: (1) a probabilistic linear framework that unifies influence of different factors contributing to the popularity of an item and inter-item competitions and (2) a more generic model, with a deep probabilistic machinery that unifies the nonlinear generative dynamics of a collection of diffusion processes, and inter-process competition.

This cross-disciplinary collection of articles highlights the bridging of a variety of scientific branches. The chapters are designed to serve as the state of the art not only for students and new entrants but also for experts who intend to pursue research in this field. All the chapters have been carefully peer-reviewed in terms of their scientific content as well as readability and self-consistency. We would like to thank the authors for their contributions and their careful consideration of the editorial comments. Moreover, we acknowledge all the reviewers as listed below for their constructive criticisms, comments, and suggestions, which have significantly improved the quality of the chapters. Also, we acknowledge Satadal Sengupta for maintaining the electronic platform for the manuscript submission, management and monitoring. Finally, we are extremely grateful to the entire support team from Springer for their help that made the timely publication of this volume possible.

Berlin, Germany	Fakhteh Ghanbarnejad
Saarbrücken, Germany	Rishiraj Saha Roy
Köln, Germany	Fariba Karimi
Louvain-la-Neuve, Belgium	Jean-Charles Delvenne
Kharagpur, India	Bivas Mitra

List of Reviewers (Alphabetically Ordered by Last Names)

Edmund Barter, Vitaly Belik, Parantapa Bhattacharya, Sanjukta Bhowmick, Tanmoy Chakraborty, Jean-Charles Delvenne, Mauro Faccin, Saptarshi Ghosh, Martin Gueuning, Jean-Loup Guillaume, Flavio Iannelli, Lucas Jeub, Suman Kalyan Maity, Alexey Medvedev, Andrew Mellor, Subrata Nandi, Camille Roth, Koustav Rudra, Michael Schaub, Sandipan Sikdar, and Simon Walk.

Contents

Part I
Network Structure

An Empirical Study of the Effect of Noise Models on Centrality Metrics

Soumya Sarkar, Abhishek Karn, Animesh Mukherjee, and Sanjukta Bhowmick

Abstract An important yet little studied problem in network analysis is the effect of the presence of errors in creating the networks. Errors can occur both due to the limitations of data collection techniques and the implicit bias during modeling the network. In both cases, they lead to changes in the network in the form of additional or missing edges, collectively termed as noise. Given that network analysis is used in many critical applications from criminal identification to targeted drug discovery, it is important to evaluate by how much the noise affects the analysis results. In this paper, we present an empirical study of how different types of noise affect real-world networks. Specifically, we apply four different noise models to a suite of nine networks, with different levels of perturbations to test how the ranking of the top-k centrality vertices changes. Our results show that deletion of edges has less effect on centrality than the addition of edges. Nevertheless, the stability of the ranking depends on all three parameters: the structure of the network, the type of noise model used, and the centrality metric to be computed. To the best of our knowledge, this is one of the first extensive studies to conduct both longitudinal (across different networks) and horizontal (across different noise models and centrality metrics) experiments to understand the effect of noise in network analysis.

Keywords Noise models in networks · Centrality metrics · Accuracy of analysis

---

S. Sarkar · A. Mukherjee
IIT Kharagpur, Kharagpur, West Bengal, India
e-mail: soumya015@cse.iitkgp.ernet.in; animeshm@cse.iitkgp.ernet.in

A. Karn
NIT Durgapur, Durgapur, West Bengal, India

S. Bhowmick (✉)
University of North Texas, Denton, TX, USA
e-mail: sanjukta.bhowmick@unt.edu

F. Ghanbarnejad et al. (eds.), *Dynamics On and Of Complex Networks III*,
Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-030-14683-2_1

1 Introduction

In recent years, network analysis has become an important mathematical tool for studying the interactions of entities in complex systems. The entities are represented as the vertices and their dyadic relations are represented as edges. The structural properties of the network provide insights to the characteristics of the underlying system. For example, high centrality vertices point to important proteins in protein–protein interaction networks [5] and groups of tightly connected vertices, or communities represent groups of friends in social networks [19].

An important yet little studied problem in network analysis is the effect of the presence of errors in creating the networks. The errors primarily occur at two stages: First, they occur when collecting real-world data. The measurements of any physical system inherently include some degree of error, which is propagated to the network model. Second, the errors occur during the creation of the network model. The interrelations (here edges) are often determined based on the subjective evaluation of the modeler. For example, in a gene correlation network, two vertices (genes) are connected by an edge only if the correlation between the two genes is higher than a specified threshold. However, due to the absence of a standard value, this threshold is decided by the person creating the model.

These errors in data collection and modeling are manifested as structural changes in the network, in the form of additional or missing edges, collectively termed as *noise*. An important question is by how much this noise affects the analysis results. In particular, since network analysis is used in many critical applications from criminal identification [12] to targeted drug discovery [3], a drastic change in the accuracy can lead to serious consequences.

Overview In this paper we present an extensive empirical study of how different types of noise affect real-world networks. Specifically, we apply four different noise models. These are random deletion of edges (edge deletion), random addition of edges (edge addition), swapping the end-points of a pair of edges (edge swap), and overlaying the network with a random graph such that only the edges present in either the original or the random network are kept (edge XOR). Each of these noise models is applied to a suite of nine networks, with different levels of perturbations to test how the ranking of the top-k centrality vertices changes.

We measure the change in the ranking using the Jaccard index (JI). We test how the following centrality metrics, *degree centrality*, *betweenness centrality*, and *closeness centrality*, are affected by these noise models. While there have been several studies [4, 6, 18] conducted on individual noise models and how they affect network properties, to the best of our knowledge, this is one of the first studies to conduct both a longitudinal (across several different networks) and a horizontal (across noise models and centrality metrics) evaluation of the effect of noise in networks. Some of the results that we observe through these experiments are:

- *Variations in Noise Models:* Edge swap produces the highest average JI, i.e., the least amount of change in the vertex ranking. This is followed by, in order of highest to lowest average JI, edge deletion, edge XOR, and finally edge addition.
- *Variations in Network Structure:* Networks from the technological domain, particularly the autonomous network AS2 and the peer-to-peer network P2P, show the most stability, i.e., high JI across all noise models. The network of the power grid of the Western United States shows the least stability with lowest JI across all noise models. The other networks show high stability for edge swap and edges deletion and low stability for the other two models.
- *Variations in Centrality Metrics:* Of the three centrality metrics, degree centrality was the most stable, and betweenness centrality was the least stable.

The remainder of this paper is organized as follows: In Sect. 2, we present our experimental methodology along with the datasets and definitions of the centrality measures. In Sect. 3, we describe the effect of the noise models and provide a summary of our observations. In Sect. 4, we discuss related research in this domain. We conclude in Sect. 5 with an overview of our future research plans.

2 Experimental Setup

In this section we provide a brief description of our experimental setup, including description of the networks used, the definitions of the centrality metrics, and an overview of how we conducted the experiments.

2.1 Test Suite of Networks

We consider the following nine networks that were collected from public repositories [8, 13]. We group these networks into three categories: (1) technological networks that are formed through interconnections of routers or peer-to-peer networks, (2) social networks that are formed through collaborations, chats, or linking between blogs, and (3) miscellaneous networks formed from other varied applications, including biological networks, software networks, and power grids. A summary of the networks, their descriptions, and sizes is given in Table 1.

2.2 Centrality Metrics

In our experiments we study the stability of the following centrality metrics. Given a graph $G(V, E)$, with $|V|$ vertices and $|E|$ edges, the metrics are computed as:

Table 1 Description of the nine networks in the test suite

Network	Description	Node	Edges
Technological networks			
AS	Network of routers obtained from University of Oregon Route Views Project	6474	13,895
CA	Network of routers obtained from Center for Internet Data Analysis (CAIDA)	16,493	66,744
P2P	Gnutella peer-to-peer file sharing network	26,518	65,369
Social networks			
APH	Collaboration network of authors of papers posted in arXiv's Astrophysics	16,046	121,251
AnyBeat	Social network where users connect anonymously	12,645	67,053
Blog	Network of front-page hyperlinks between blogs related to the 2004 US election	1224	19,025
Miscellaneous networks			
PW	Network of power grid of the Western United States	4941	6594
BIO	Protein–protein interaction network	7393	25,569
SW	Dependency network of classes in JUNG and javax	6120	50,535

From left to right, the columns are: abbreviation of the network name, short description of the network, the number of vertices, the number of edges, and the global clustering co-efficient

Degree centrality, $D(v)$, of a vertex v measures the number of its neighbors. *Closeness centrality* of a vertex v measures its average distance from all other vertices in the network. It is computed as $CC(v) = \frac{|V|}{\sum\limits_{s \neq v \in V} dis(v,s)}$, where $dis(v, s)$ is the length of a shortest path between v and s.

Betweenness centrality of a vertex v is defined as the fraction of the total number of shortest paths that pass through the vertex. It is computed as $BC(v) = \sum\limits_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$, where σ_{st} is the total number of shortest paths between s and t, and $\sigma_{st}(v)$ is the total number of shortest paths between s and t that pass through v.

2.3 Methodology

To test the stability of the networks and centrality metrics under different noise models, we perform the following experiment. We first compute the centrality values of each network, and rank the vertices from highest to lowest centrality values. There can be potentially different ranking for each centrality metric. We test the effect on rank, rather than the value of the centrality, because most applications, such as information spreading or vaccination, require finding the high ranked vertices and do not require their exact value. We apply the noise models to the networks over different levels of perturbations, and compute the centrality, and subsequent ranking on the perturbed network.

We then use Jaccard index (JI) to measure how many of the top-k vertices from the original ranking are retained. The JI of two sets A and B is given by $\frac{A \cap B}{A \cup B}$. The highest value is 1, when both sets A and B are the same and the lowest value is 0, when the sets A and B do not have any common elements. We test JI for the top ranked 5, 10, 25, and 50 vertices. Since the noise models are stochastic, test is repeated 5 times, and the average JI over the five tests is reported for each tuple of noise model, network types, centrality metric, and perturbation level.

3 Empirical Results

We describe the stability results as per our experiments on the four noise models. We test one model for only addition of edges, one for only deletion, and two models that involve both addition and deletion of edges. Addition and deletion are the units of change in a network, so we study them individually. Variations of the other two models have been used in [2] and [6]. We present the detailed results for each model separately in Sects. 3.1–3.4 and summarize our findings from all these experiments in Table 2.

3.1 Edge Addition

In this noise model we add edges to the network. We select a pair of vertices from the set V with probability $\frac{p}{|V|}$. If the edge is not already part of the network we add it to the network. Figures 1, 2, and 3 show how the top-k centralities change as the value of p is increased. The values of p ranged from 0.5, 1.5, 2.5, 3.5, and 4.5. As

Table 2 Average across centrality metrics for each noise model for each network

Network	Edge deletion	Edge addition	Edge swap	Edge XOR
Technological networks				
AS2	0.8	0.59	0.9	1
CA	0.7	0	0.96	0
P2P	0.68	0.61	1	0.80
Social networks				
AnyBeat	0.8	0	1	0
APH	0.6	0	1	0
Blog	0.97	0.09	0.93	0.03
Miscellaneous networks				
BIO	0.78	0	0.88	0
PW	0.47	0.02	0.34	0.11
SW	0.95	0.13	0.9	0.21

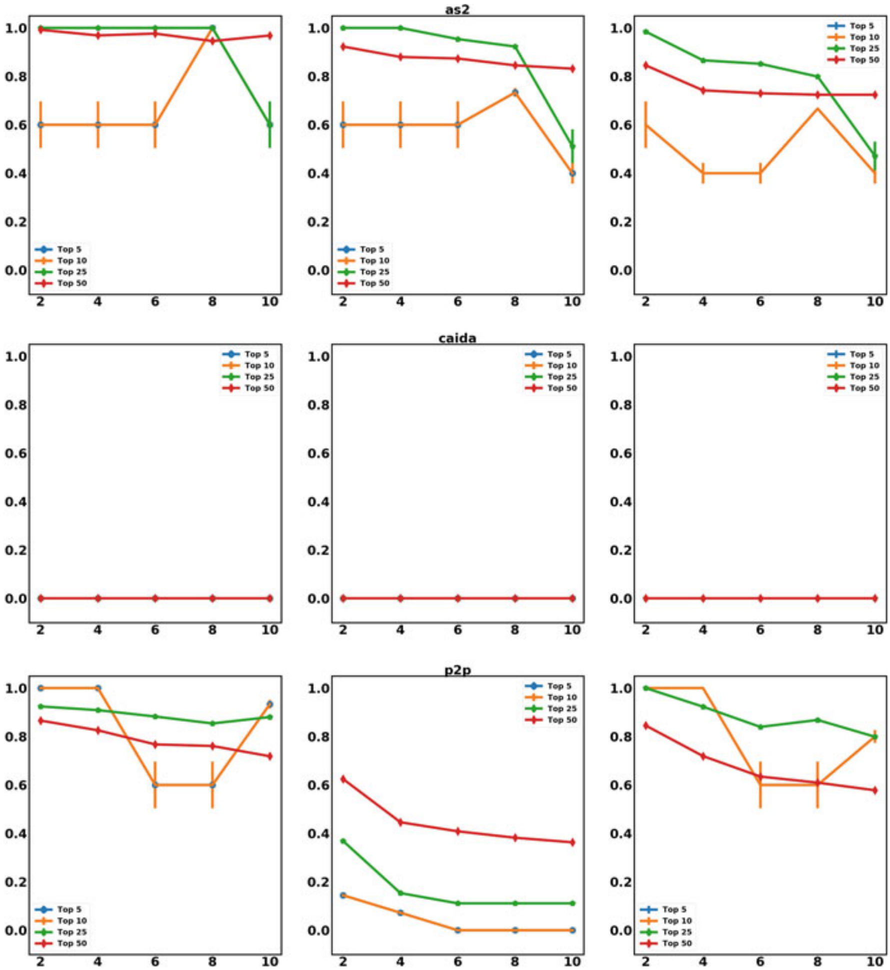

Fig. 1 Effect of edge addition on technological networks. X-axis: perturbations values; Y-axis: Jaccard index. Left: degree; middle: betweenness; right: closeness. Top to bottom: AS, CAIDA, and P2P

can be seen, apart from AS2 and P2P, all the other classes of networks exhibit very low JI for every perturbation and every value of k. Thus *edge addition even at small levels of perturbation can significantly change the ranking*.

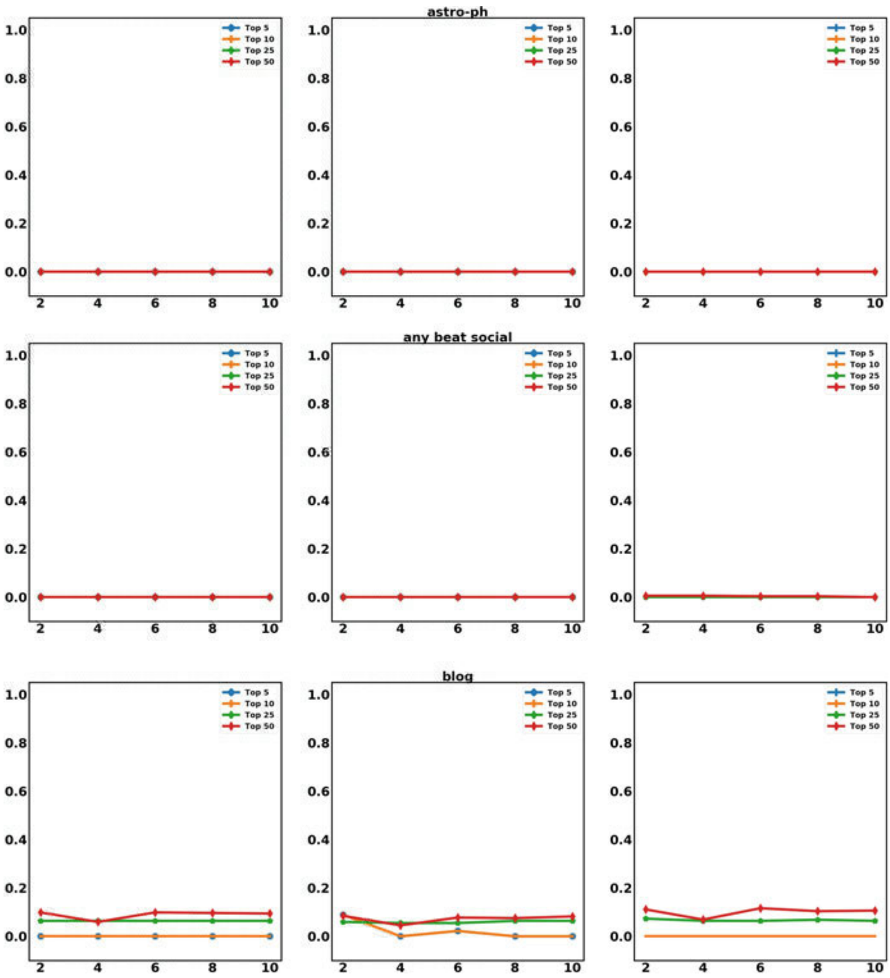

Fig. 2 Effect of edge addition on social networks. X-axis: perturbations values; Y-axis: Jaccard index. Left: degree; middle: betweenness; right: closeness. Top to bottom: APH, AnyBeat, and Blog

3.2 *Edge Deletion*

In this noise model we delete edges from the network. We select an existing edge from the set E and remove it with a probability of p. In our experiments we set p to range from 2, 4, 6, 8, and 10% of the edges. Figures 4, 5, and 6 show how the top-k centralities change as the value of p is increased.

We observe that in contrast to the edge addition model, the JI values for degree and closeness centralities are generally high for all networks. The behavior of the JI values of betweenness centrality varies from being 1, i.e., ranking unaffected (Blog)

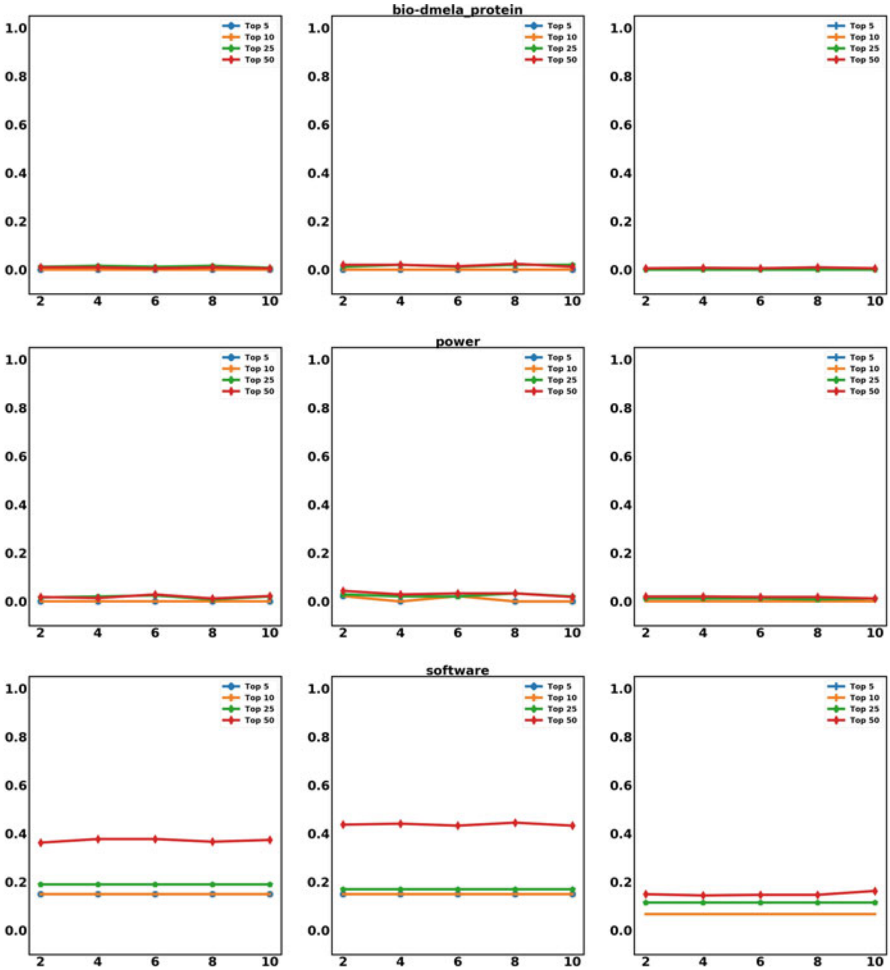

Fig. 3 Effect of edge addition on miscellaneous networks. X-axis: perturbations values; Y-axis: Jaccard index. Left: degree; middle: betweenness; right: closeness. Top to bottom: BIO, PW, and SW

to gradually decreasing (AnyBeat) to being 0, i.e., ranking completely changed (APH). We conclude that *uniform deletion does not significantly affect the ranking of degree and closeness centralities.*

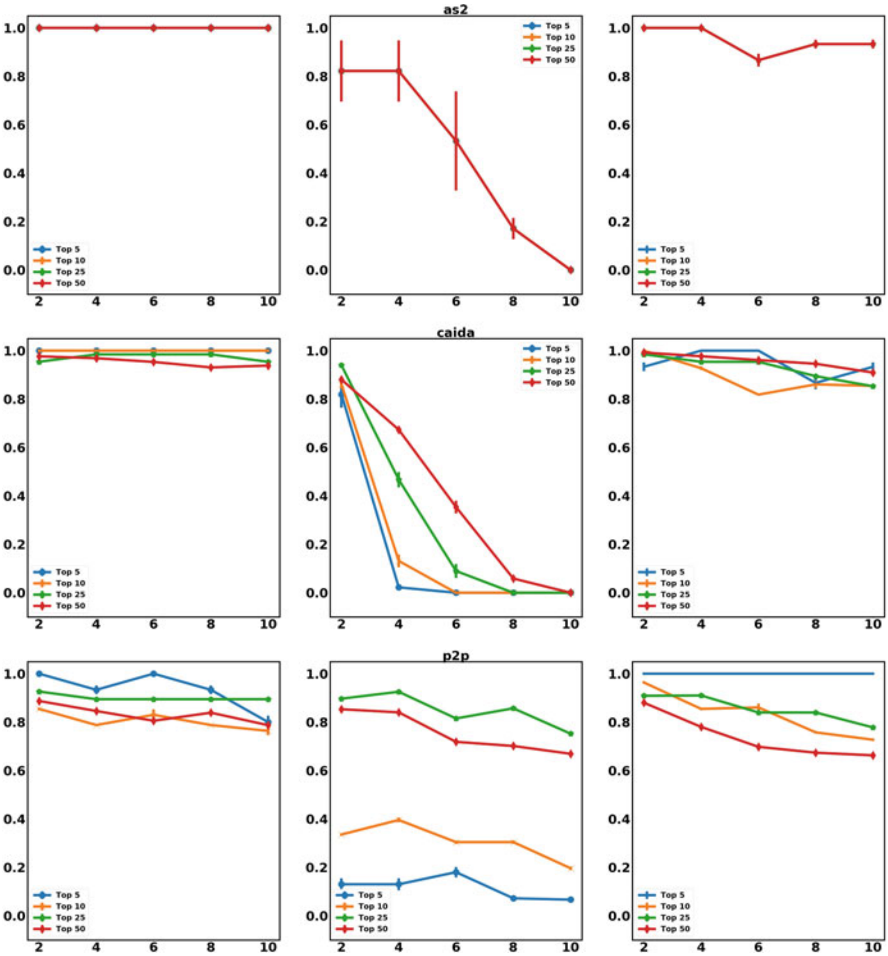

Fig. 4 Effect of edge deletion on technological networks. X-axis: perturbations values; Y-axis: Jaccard index. Left: degree; middle: betweenness; right: closeness. Top to bottom: AS, CAIDA, and P2P

3.3 Edge Swap

We now consider noise models that include both addition and deletion of edges. The first example we consider is swapping edges between two pairs of connected vertices. Let us consider two edges, (a, b) and (c, d), where none of the vertices a and b are connected to vertices c and d. In the swapping model, we disconnect the edge between a and b, and between c and d. Then, to maintain the degree of the vertices, we reconnect a with c and b with d. A version of this noise model was used in [6] to measure robustness of communities.

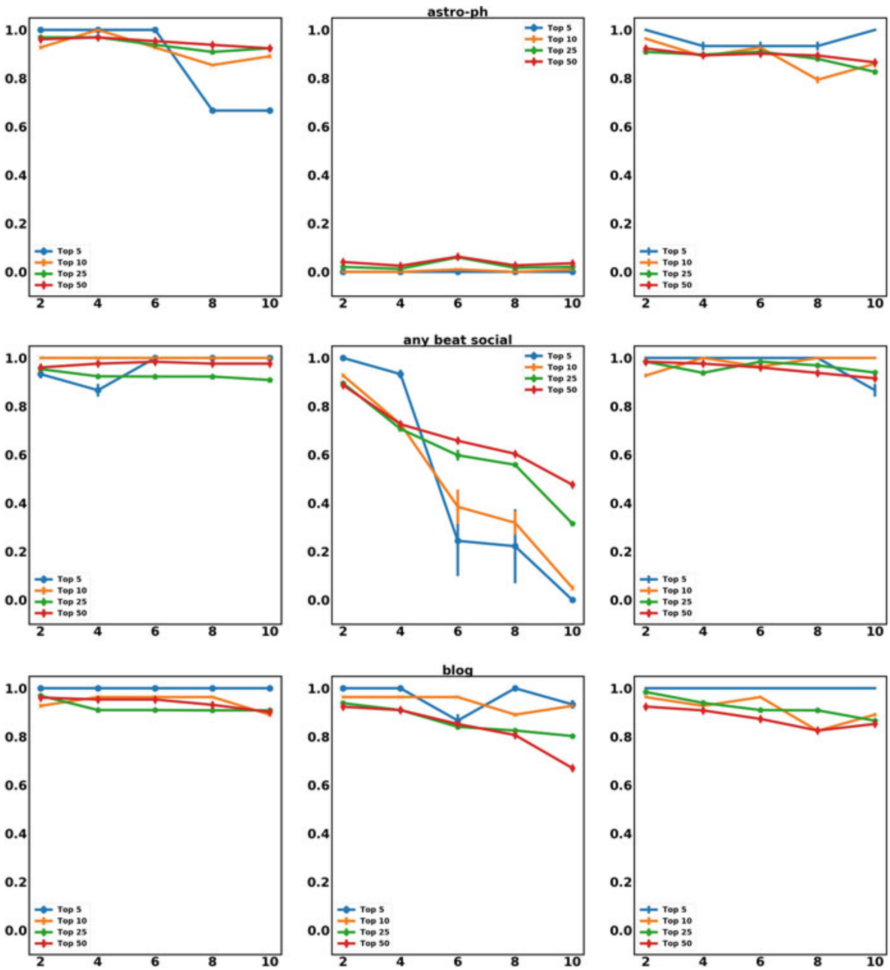

Fig. 5 Effect of edge deletion on social networks. X-axis: perturbations values; Y-axis: Jaccard index. Left: degree; middle: betweenness; right: closeness. Top to bottom: APH, AnyBeat, and Blog

Figures 7, 8, and 9 show how the top-k ranking changes as the value of p, the percentage of edge selected, is increased. Values of p are 2, 4, 6, 8,and 1. Due to the characteristics of the model, the ranking of degree will remain mostly unchanged. We see that *for most networks and centrality metrics, the JI values are high indicating that swapping does not significantly perturb the centrality ranking.* The exceptions are the power network, and some cases of betweenness centrality.

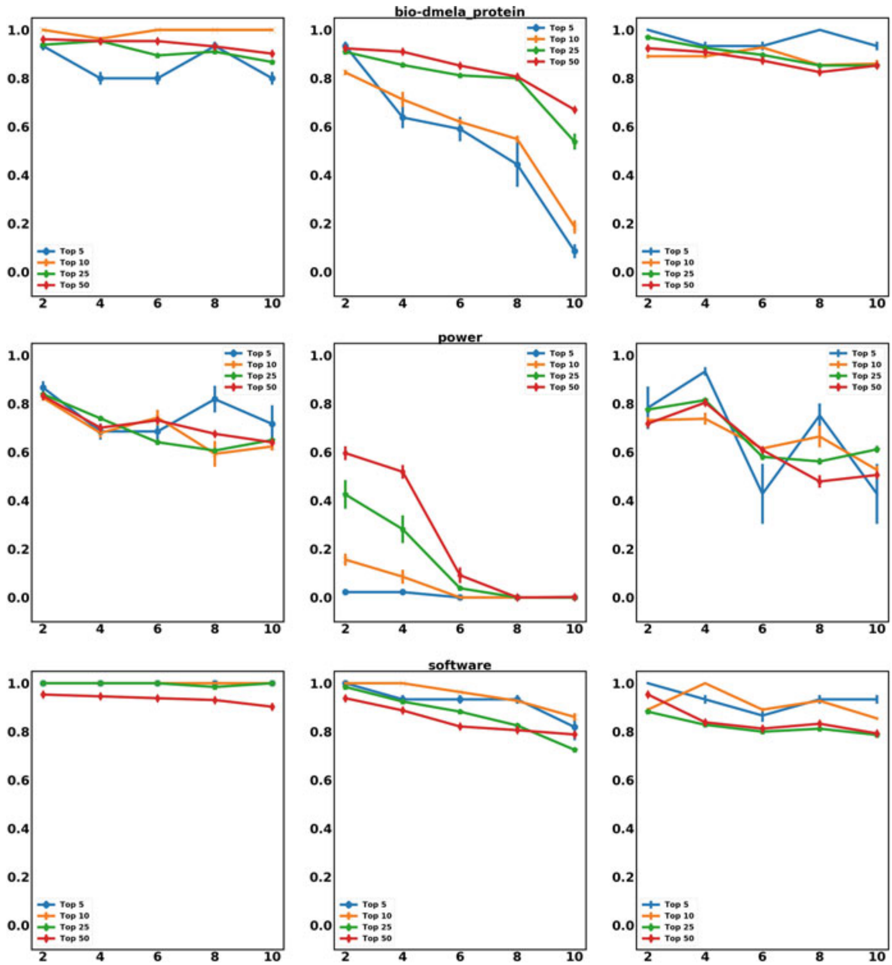

Fig. 6 Effect of edge deletion on miscellaneous networks. X-axis: perturbations values; Y-axis: Jaccard index. Left: degree; middle: betweenness; right: closeness. Top to bottom: BIO, PW, and SW

3.4 Edge XOR

In this model, we also consider both addition and deletion of edges. Here we create a random graph $R(n, p)$ with the same number of vertices, $|V| = n$ as the original graph G. We perform a perturbation in the form that if an edge (a, b) is present in both R and G, the edge is deleted from the G. However, if the edge is present in R but not in G, we add the edge to G. We term this as the XOR perturbation, because of its similarity to the boolean XOR operation (output is true if exactly one, but not both conditions are true). A version of this model was used in [1], for testing

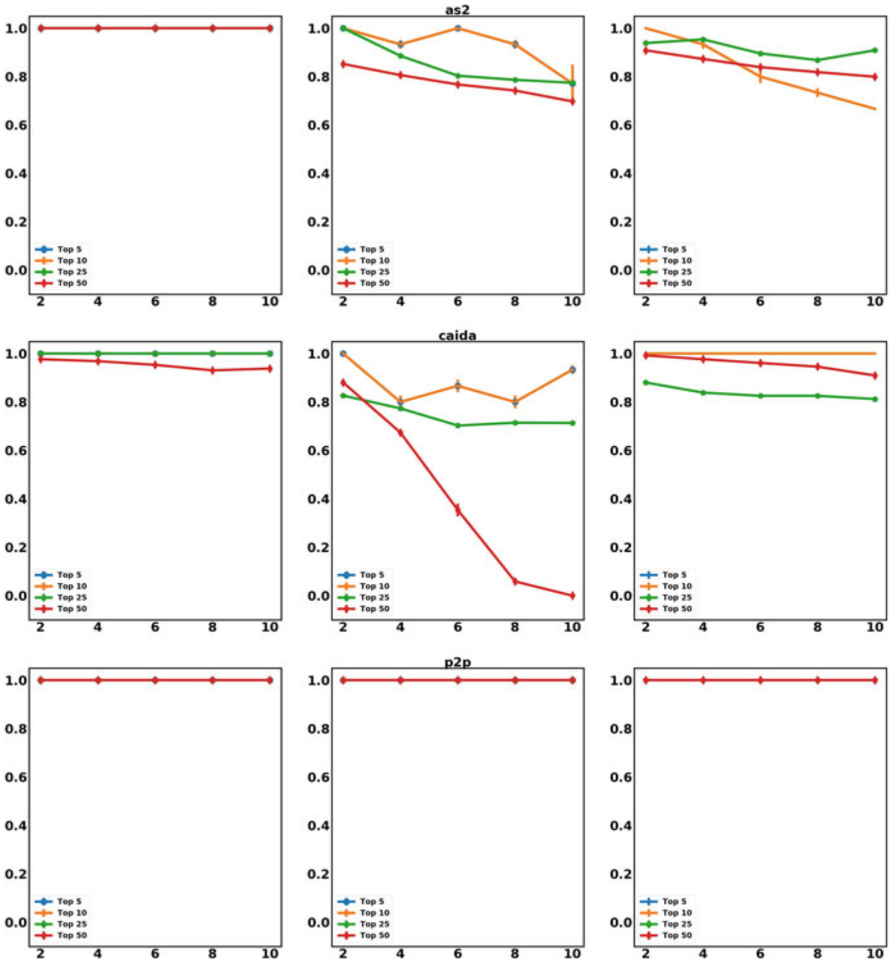

Fig. 7 Effect of edge swapping on technological networks. X-axis: perturbations values; Y-axis: Jaccard index. Left: degree; middle: betweenness; right: closeness. Top to bottom: AS, CAIDA, and P2P

robustness of the k-core. Because the networks are sparse, more edges will be added than deleted in the perturbed network. The probability with which the edges in the random network were connected is varied from 0.5, 1.5, 2.5, 3.5, and 4.5.

The results in Figs. 10, 11, and 12 show that for all networks, except for AS2 and P2P, the JI value is close to zero for all networks and centrality measures. This indicates that the *XOR model can easily disrupt the ranking of the high centrality vertices*, even for low perturbations.

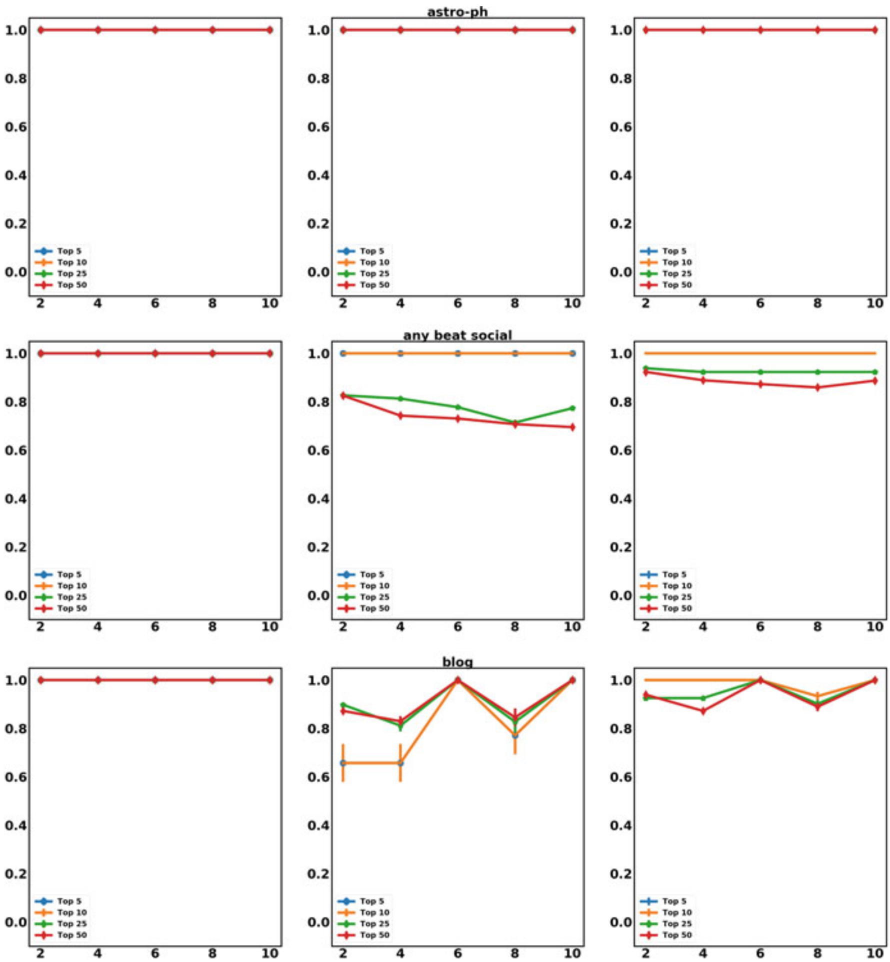

Fig. 8 Effect of edge swapping on social networks. X-axis: perturbations values; Y-axis: Jaccard index. Left: degree; middle: betweenness; right: closeness. Top to bottom: APH, AnyBeat, and Blog

3.5 Summary of the Results

In this set of experiments, we studied how different noise models affect the ranking of high centrality vertices. Table 2 summarizes the average JI for each network over all the centrality metrics, and for each noise model.

From the results it can be clearly seen that the edge deletion and edge swap affect the ranking of the high centrality vertices far less than the edge addition and edge XOR. Out of these two, the edge addition model is more disruptive. Note that almost any new edges has the potential to change the route of the shortest paths, leading