Journal of Business Research xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

Ambiguity of network outcomes

Matteo Cinelli

Applico Lab, CNR-ISC, Uos "Sapienza", 00185 Rome, Italy

ARTICLE INFO

Keywords:
Ambiguity
Network management
Complexity
Analytics

ABSTRACT

The extent to which available data is continuously growing in terms of volume is forcing organizations to contend with and seek to resolve the so-called Big Data Challenge. Big data comes or can be structured in the form of networks from which information can be extracted via statistical and computational tools. The results of such investigations can be generally referred to as network outcomes. Such outcomes, despite being often characterized by a inner ambiguity, need to be well understood and interpreted in order to exploit the potentialities of network data, especially in practical situations. For this reason, addressing the ambiguity of network outcomes becomes a key issue in business-related environments, where the possibility of rapidly interpreting and properly exploiting network data can positively affect performances. In this paper, we propose a framework to face ambiguity of network outcomes that, by means of specific solutions, allows practitioners to successfully interpret and exploit the obtained outcomes.

1. Introduction

Ambiguity – the quality of being open to more than one interpretation – is a source of complexity. It strongly affects the way in which we make decisions regime uncertainty as well as the confidence we have in the outcome of such decisions. Ambiguity derives from having a limited knowledge of the process that generated the outcome (Einhorn & Hogarth, 1985) thus many real-world systems, of which knowledge is indeed limited, can be deemed ambiguous in some of their aspects. Building on the fact that ambiguity affects various elements of organizational settings, numerous scholars have approached such an issue through the lens of business practices and management. For instance, the role of ambiguity has been linked to the efficiency of the decision-making process (Einhorn & Hogarth, 1986; Frisch & Baron, 1988), to the conflict among roles within organizations (Kahn, Wolfe, Quinn, Snoek, & Rosenthal, 1964) and to the effectiveness of leadership (Pfeffer, 1977). The concept of ambiguity has also been studied in conjunction with business performance (Powell, Lovallo, & Caringal, 2006), taking into account the role of causal ambiguity; the condition under which neither the firm nor its rivals can determine the causes of firm performance. More recently, research on the role of ambiguity has shifted towards problems related to uncertainty in financial systems (Epstein & Schneider, 2008). In light of the benefits that insights from complexity theories can bring to organizations (Burnes, 2005), this paper explores the role of ambiguity from a complex systems perspective by providing practical solutions to such issues commonly related to the concept of ambiguity.

Considering a complex system as the outcome of a generative process, of which in general we know only a limited number of drivers, it is reasonable to suggest that complex systems display, through their very nature, a certain level of ambiguity. A well-known example of a complex system is represented by networks through which we are able to schematically represent and model a vast portion of real-world systems (Newman, 2010). The usefulness of networks has been proved in numerous different domains, including business-related environments (Håkansson & Ford, 2002), and there has been extensive research into networks of social relations and industrial relations, as well as networks representing relationships of a more technical nature.

Networks are made up of nodes interacting with one another through links that, realizing different patterns of connections, turn networks into relatively complex objects to analyze. The complexity of networks is enhanced by an increasingly simplified and streamlined way to collect, process and store large amount of data. Furthermore, the fact that current available networks are continuously growing – e.g., in terms of size or temporal resolution – has caused many to adopt the term big data when describing them. As such, understanding, exploiting and integrating complex networks within tools of business intelligence and analytics is pertinent for those companies and organizations hoping to resolve the challenges posed by big data (Sivarajah, Kamal, Irani, & Weerakkody, 2017); a process that requires a wide body of knowledge, skills and expertise that is still under development in the field of complex networks. Indeed, being that recent networks orders of magnitude are larger than those available a few years ago (Wasserman & Faust, 1994), the classic tools conceived for social network analysis

E-mail address: matteo.cinelli@roma1.infn.it.

https://doi.org/10.1016/j.jbusres.2020.02.031
Received 15 March 2019; Received in revised form 12 February 2020; Accepted 13 February 2020
0148-2963/

suffer from computational issues and of lack of explanatory power. Furthermore it is generally impossible to obtain information from a network by eyesight alone or to compute network measures by hand. For this reason, the investigation of complex networks is conducted via specific measures that, acting as summary statistics, are helpful in supporting decisions related to networked systems. Any measure that derives from the network topology, including distributions, summary statistics and results of algorithms, can be considered and is here referred to as network outcome.

Considering that networks, being complex systems, are characterized by a certain level of ambiguity, in this paper we argue that such ambiguity can be inherited by network outcomes causing issues of interpretability even when we have complete knowledge about the system that generated the outcome. Indeed, when referring to network outcomes, ambiguity has to be considered as a problem of unequivocal interpretability of the obtained outcome in relation to the network itself. Such a possibility for interpretative spaces caused by ambiguity may entail serious issues for companies and organizations that are ruled by the "management by measurement" practice (Noordegraaf & Abma, 2003) as well as for those who adopt performance-oriented decisions. Considering that networks are now a very popular, required and available scheme of representation and analysis, it becomes crucial to point out the risks of ignoring inner ambiguity when making decisions based on measurement and analytics.

Discussing the ambiguity of network outcomes aims at raising awareness about the potential flaws and false positives related to such popular tools and to provide innovative solutions.

For instance, a company offering a vehicle- sharing service may easily profile its customers (or discover criticalities about its vehicles) by building different kinds of networks using data from social media or by exploiting data about geographical proximity of customers. In such a context, while using network outcomes as an advising tool for supporting decisions is common practice, ignoring their limitations and potential ambiguity may result in incorrect interpretation of the data and consequently in wrong decisions. This aspect raises an important issue especially for relatively vulnerable companies such as start-ups, where the exploitation of data-driven solutions and business intelligence solutions is more common.

The paper is structured as follows: Section 2 discusses the concept of ambiguity related to different types of network outcomes. Section 3 presents the data. Section 4 quantitatively discusses the issues related to the ambiguity of network outcomes at node level and presents a method for addressing such ambiguity. Section 5 quantitatively discusses the issues related to the ambiguity of network outcomes at network level and shows how embedding external information in the structure of the network can provide a clearer interpretation of the observed outcome. Finally, Section 6 presents conclusions and directions for future works.

2. Ambiguity of network outcomes

Network outcomes can be considered at different levels and they can provide information either about the single element (node-level measures) or about the whole system (network-level measures). Both kinds of measures, due to different reasons, may be deemed hard to interpret (ambiguous) and, if so, may negatively affect the process of decision making and the management of the considered system. Indeed, in the case of large networks, node-level measures may be hard to interpret because of the presence of thousands of nodes or because of the relatively small difference between two nodes displaying different values of the considered measure. In order to address this issue, a common practice inherited from the study of relatively small networks in the field of social networks analysis (Wasserman & Faust, 1994), is to consider only a smaller set of values of node-level measures by means of rankings. This practice limits the analysis to an arbitrarily small fraction of the nodes (commonly top 10, 20, 100), dropping important pieces of information. Another common practice is to resume node-level measures into distributions that, despite their undeniable usefulness, inevitably provide aggregate information (Newman, 2010). These kind of practices related to node-level measures may introduce arbitrariness in terms of interpretation and create a labile ground for decision making by introducing arbitrariness within a system already characterized by inner ambiguity.

Differently from the case of node-level outcomes, the problems related to the interpretability of network-level outcomes are somewhat more complex to tackle. When we take into account such kind of measures we are basically resuming the whole network by means of a quantity that only partially explains the underlying structure of interconnections. Such a problem is typical in statistics when, for example, we observe the same value of the correlation coefficient for very different distributions of the data (Anscombe, 1973). As explained in Peel, Delvenne, and Lambiotte, 2018, the exact same problem can be observed in the case of network-level measures such as the assortativity coefficient (Newman, 2003). Additionally, certain-network level measures (e.g., modularity that is a quality function associated to the process of network clustering) are outcomes of optimization processes and thus suffer from algorithm dependency (Fortunato & Hric, 2016) or, as explained in Good, De Montjoye, and Clauset, 2010, display different configurations that provide the same or very similar outcomes (i.e., degenerate and near-degenerate solutions). Finally, network-level measures may present further problems of interpretability, by displaying constraints determined by the "rigidity" of the underlying structure of conncetions that make them hard to compare against one other (Cinelli, Ferraro, & Iovanella, 2017; Cinelli, Peel, Iovanella, & Delvenne, 2019). Therefore, when supporting decision making, network-level measures also raise issues of interpretability, resulting in potential decisions that may be far from displaying the expected effects.

In order to ease the process of decision making and to provide practitioners with tools that can help in interpreting network outcomes, we propose different solutions applicable in the case of both node-level and network-level outcomes. In the case of the former, we aim to resolve the issue of hardness of interpretation by adopting a non-parametric method capable of categorizing classes of important nodes and thus reducing the arbitrariness of commonly used rankings. The division into a relatively small number of categories is important to support decisions, particularly when considering the presence of certain humans' cognitive limitations which justify the use of psychometric-aware scales such as the Likert scale (Likert, 1932).

In the case of the latter, we use a computed network outcome in combination with other decision criteria capable of either validating or imposing certain confidence intervals on the considered network outcome. In other words, we aim to evaluate the network outcome in combination with external information, considered of interest for the problem, in order to understand the extent to which the observed structure matches other relevant data or if such external data are irrelevant to the network structure. Such approach to the evaluation of network-based outcome would help in addressing ambiguity by means of results characterized by a higher practical validity.

3. Data description

In order to clarify the importance of the proposed approach we will consider, as an example, a social network that will be used throughout the paper. The network taken into account is an online social network of students from the Harvard College extracted from the Facebook100 dataset (Traud, Mucha, & Porter, 2012). The Facebook100 dataset contains an anonymized snapshot of the friendship connections among users affiliated with the first 100 colleges admitted to Facebook. The obtained friendship network, depicted in Fig. 1, is related to Harvard College and its largest connected component comprises nodes and $m = 682824$ links.

Fig. 1. Representation of the Harvard College social network according to the Facebook100 dataset. Nodes are colored according to the community partitioning obtained using the Louvain method.

4. Node-level outcomes

As discussed in Section 1, node-level outcomes can be affected by a certain level of ambiguity related to the difficulty in their interpretation. Such outcomes, also called centrality measures, can be of different types and they are able to quantify the importance of nodes from different perspectives. For instance, there are measures able to quantify the importance of the nodes based on the number and on the quality of their connections (e.g., degree and eigenvector centrality), measures that consider a node important if, being present in many paths between couples of other nodes, it is involved in the process of information exchange (e.g., betweenness and closeness centrality) and measures that quantify the importance of a node relatively to the characteristics of its neighbourhood (e.g., local clustering coefficient and local efficiency). It is also important to recall that such node-level measures are sometimes network-level measures that are properly split among the network nodes. Considering the Harvard friendship network of the Facebook100 dataset, we will compute two node-level measures, namely degree $\mathbf{k}$ and local clustering coefficient $\mathbf{c}$. The node degree k_i measures the number of connections of node i and it can be written, using the adjacency matrix $\mathbf{A}$, as:

$$k_i = \sum_{j=1}^{n} A_{ij}. \tag{1}$$

The matrix A is a binary and symmetric matrix in which the element $A_{ij} = 1$ ($A_{ij} = 0$) in position ij, indicates the presence (absence) of a connection between nodes i and j.

The local clustering coefficient $c_i \in [0, 1]$ measures the density of connections of the neighbors of a given node i and it can be written as:

$$c_i = \frac{t_i}{\binom{k_i}{2}} \tag{2}$$

where t_i is the number of links between the neighbors of i. The local clustering coefficient c_i has an inherent interest since it captures the capacity of link creations among neighbors, i.e., the tendency in the network to create stable groups in the network.

A viable option to avoid arbitrary cuts of the ranking deriving from the computation of a certain centrality measure is that of dividing nodes into classes using a more principled method. In order to classify the nodes we use a non-parametric method that has been developed in the field of bibliometrics (i.e., the study of scientific productivity of academics) to classify different categories of academics by partitioning the distribution of their number of publications. The method, introduced in Schubert, Glänzel, and Braun, 1987, is referred to as Characteristics Scores and Scales (CSS) and can be consistently applied for partitioning distributions not necessarily related to the number of publications. The CSS method has been mostly exploited in the field of bibliometrics where quantities such as authors' productivity or number of citations display strongly skewed as well as heavy tailed distributions. Such heterogeneous distributions characterize also network outcomes, thus the CSS method is apt for the considered case. The technique involves reiterated truncation of a frequency distribution according to mean values μ, also known as characteristic scores. After truncating the overall distribution at its mean value, the mean of the sub-population above the first mean is recalculated; the sub-population is again truncated, and so on until the procedure is stopped (Abramo, D'Angelo, & Soldatenkova, 2017). By applying the CSS method to up to three characteristic scores while considering a node-level measure as a prominence indicator, the following five categories of nodes can be obtained:

- Zero prominence (ZP): $C = \mu_0 = 0$
- Low prominence (LP): $\mu_0 < C \leqslant \mu_1$
- Fair prominence (FP): $\mu_1 < C \leqslant \mu_2$
- High prominence (HP): $\mu_2 < C \leqslant \mu_3$
- Very high prominence (VHP): $C > \mu_3$

Fig. 2 shows the distributions of degree (number of friends of the node) and local clustering coefficient (interconnectedness of the node's friends) together with the results deriving from the application of the CSS method to such distributions of node-level outcomes.

We observe that the distributions of the degree and of the local clustering coefficient display very different shapes: the former is a heavy-tailed distribution, a common feature of sparse real networks; the second is a Gaussian-like distribution displaying right skewness. The CSS method is able to partition the two distributions in different classes regardless the distribution type, also giving us information about the proportion of the nodes in each class given their prominence with respect to a certain centrality index. In both cases we note that the majority of nodes have a centrality value below the first mean while classes related to higher performances are progressively less populated. This observation is more evident for the clustering coefficient whose distribution displays right skewness rather than a heavy tail. Such kind of partitioning of the network nodes is helpful when we are interested in discovering the elements with the highest prominence, without choosing them by means of an arbitrary cut of their ranking such as top 10 or top 20.

The CSS method is useful also in other situations where we consider those network-level outcomes whose value can be split at node level but that, despite this feature, may result in being hard to interpret. A well-known example of such kind of measure is the assortativity coefficient to degree, $r \in [-1;1]$, which measures the extent to which nodes with similar number of connections are interconnected (Newman, 2003).

The assortativity coefficient to degree can be written in the following way:

$$r = \frac{\sum_{ij} (A_{ij} - k_i k_j/2m) k_i k_j}{\sum_{ij} (k_i \delta_{ij} - k_i k_j/2m) k_i k_j} \tag{3}$$

where δ_{ij} is the Kronecker delta function. When we observe positive assortativity it means that nodes with similar degree values are more interconnected than expected at random. When we observe negative assortativity (disassortativity) it means that nodes with different degree values tend to be more interconnected than expected at random. The

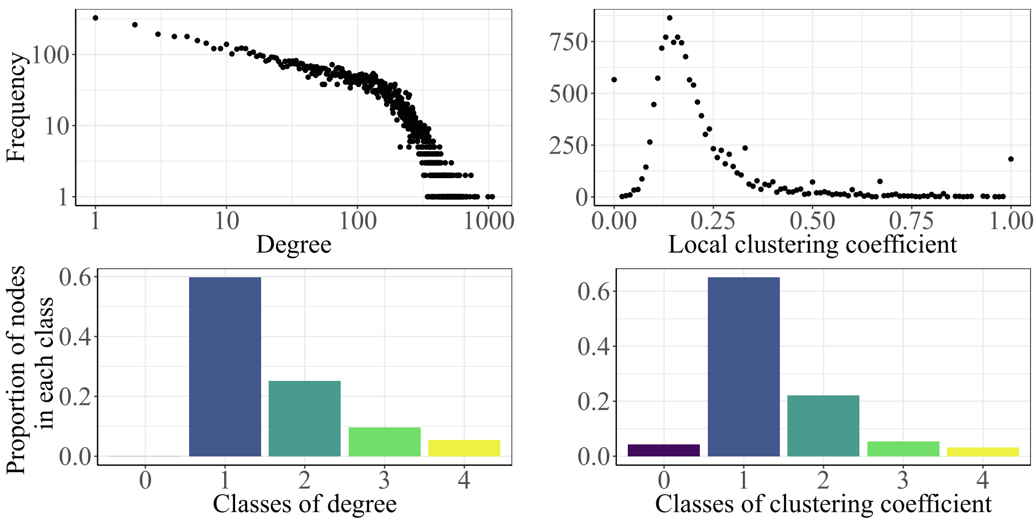

Fig. 2. Distribution of node-level outcomes and results of the CSS method. The panel in the top-left reports the degree distribution, which displays a heavy tail. The panel in the top-right reports the distribution of the local clustering coefficient, which displays a Gaussian-like distribution with right skewness. The panel in the bottom-left reports the results of the CSS method for the degree; the values of μ are: $\mu_0 = 0$, $\mu_1 = 105$, $\mu_2 = 202$, $\mu_3 = 291$. The panel in the bottom-right distribution reports the results of the CSS method for the local clustering coefficient; the values of μ are: $\mu_0 = 0$, $\mu_1 = 0.21$, $\mu_2 = 0.37$, $\mu_3 = 0.61$.

assortativity coefficient is a very important measure for the study of networks (as much as Pearson's correlation is important in statistics) but it is often summarized with the fact that high-degree nodes are connected to high-degree nodes in the assortative case, and low-degree nodes are connected to low-degree nodes in the disassortative one (Newman, 2010). This simplification of the information content related to assortativity may be due to the fact that the overall concept of association between structurally similar nodes may result hard to interpret and ambiguous, especially if the features of the underlying degree distribution are not known. The assortativity coefficient to degree is a network-level outcome but it can also be expressed considering the contribution of each node to the global coefficient r in a way such that $\sum_i r_i = r$ (Piraveenan, Prokopenko, & Zomaya, 2008, 2010)[1].

As such, we can exploit the classes of degree obtained with the CSS method so as to avoid processing information in an either too aggregate (network-level) or too distributed (node-level) fashion, thus understanding the extent to which each class of nodes contributes to the overall degree assortativity. This procedure would be especially helpful for understanding and properly exploiting the information carried by degree assortativity under different perspectives.

The Harvard social network that we take into account displays positive assortativity $r = 0.14$ meaning that nodes with similar degree value tend to be interconnected. The global assortativity coefficient can be split into its nodal components, thus producing the scatter plot reported in Fig. 3 (*left*). In such a case we observe that most of the nodes have a roughly zero contribution to global assortativity while the nodal contribution deviates from zero for those displaying a higher degree. Considering in more detail the contribution to assortativity by classes of degree, obtained via the CSS method and displayed in Fig. 3 (*right*), we note that positive assortativity is mostly due to links among high-degree nodes rather than to links among low-degree ones. Such tight association between high-degree nodes (i.e., those displaying very high prominence in terms of number of connections) is explanatory since it implies that the observed positive degree assortativity is not due to the connections among similar nodes but rather to the connections among highest-degree nodes. Such an observation has relevant implications in practice. Suppose that we aim at inferring information about the users of a social network by exploiting their tendency to associate with similar ones; by looking at the assortativity coefficient we would say that the general trend is towards association among users with similar degree. However, such a trend is actually misleading since it is induced by a strong localization of connections among nodes displaying a high prominence in terms of degree. Once it is discovered that mostly high-degree nodes are very well interconnected (i.e., the network displays a rich-club (Zhou & Mondragón, 2004; Cinelli, 2019)), one could leverage such knowledge to optimize the diffusion of information about a specific product or service. Using the insights obtained resolving the ambiguity of the assortativity coefficient, we could exploit the density of connection among hubs being almost sure that targeting even just one hub would cause the complete diffusion of information across the whole network.

5. Network-level outcomes

Network-level outcomes are used to summarize information about the whole network in a unique value. Such an aspect is a source of ambiguity since we may ignore further information about the network and rely only on a single measure to interpret the data. In Section 4, we have seen that it is possible to address the ambiguity related to network-level outcomes when we are allowed to obtain their value split across the nodes. However, such a possibility is not valid for all the network measures and, in some cases, a better interpretation of the measure itself would not be enough to get the multiple facets of the problem being addressed. The analysis of network-level outcomes implemented in Section 5 can be complemented considering the following instance.

Suppose that a service provider wants to use information deriving from a social media to profile users in order to offer them the most suitable service. A good approach to this problem, also implemented in many recommendation systems (Lü et al., 2012), would be to retrieve communities of users and propose them a specific service by community. Indeed, it is known that communities in social networks are often made up of individuals similar in many respects, meaning the likelihood of observing similar preferences within a community is expected to be higher than across communities. Considering only the structural point of view of network data may be quite restrictive, despite being useful, when we are interested in the preferences of the users. Additionally, from the methodological point of view, it is worth noting that the task of partitioning the network into communities displays an inner ambiguity; indeed, several valid ways in which a network can be partitioned into communities exist and they are not necessarily mutually exclusive. In order to face the ambiguity deriving from the possibility of partitioning the network in different ways, it is possible to introduce a further criterion of choice by exploiting additional information about the network nodes. Indeed, users normally display additional features (sometimes called metadata) that are external to the structure of the network and that can possibly display a certain

[1] In the case of nodal assortativity to discrete metadata, refer to Peel et al., 2018

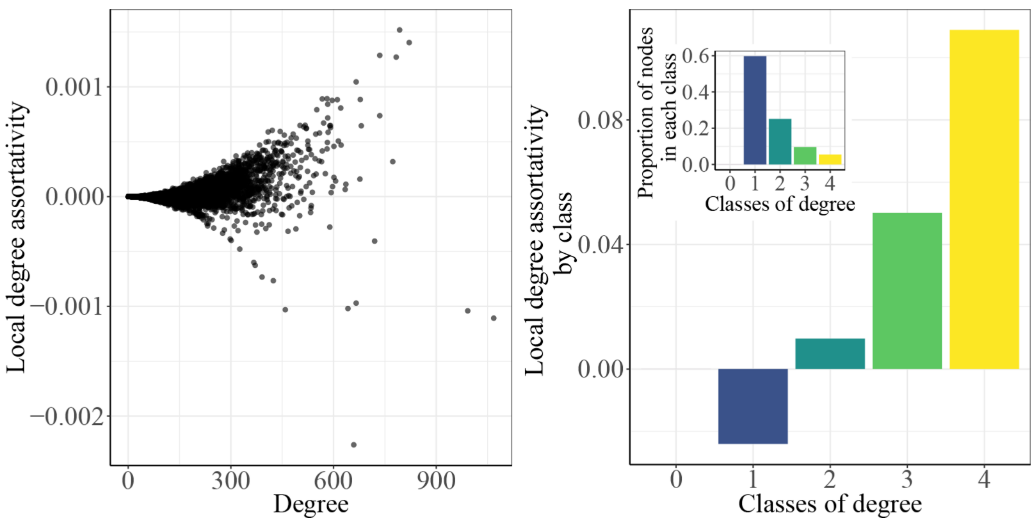

Fig. 3. Scatter plot of local assortativity to degree and results of the CSS method for local assortativity and degree. The left panel displays the values of local assortativity. Low-degree nodes are associated to values of local assortativity that are nearly zero while high-degree nodes are associated to values of assortativity that are higher in absolute value. The right panel displays the results of the CSS applied to local assortativity. The values of μ for the different classes of degree are $\mu_0 = 0$, $\mu_1 = 105$, $\mu_2 = 202$, $\mu_3 = 291$ and the corresponding overall values of local assortativity are $r(\mu_0) = 0$, $r(\mu_1) = -0.02$, $r(\mu_2) = 0.01$, $r(\mu_3) = 0.05$, $r(\mu_4) = 0.1$ The sum of the latter set of values equals the coefficient of assortativity to degree $r = 0.14$. The inset of the right panel reports the results of the CSS applied to degree.

correlation with the structure. Such metadata, in the case of a social network, can be of various types such as age, gender or preference for specific products/services. Metadata can be either uncorrelated to the network structure (providing information unrelated to the system) or correlated to the network structure (well representing the observed patterns of connections). The analysis of metadata is of extreme importance (Peel, Larremore, & Clauset, 2017; Cinelli, 2019) since, when combined with structural outcomes, it is able to ease the interpretation of the observed structural patterns thus allowing room for a consistent use of information deriving from the network structure.

In order to present the problem of interrelation between the network structure and the node metadata we will take into account the aforementioned example of a service provider willing to profile users for commercial purposes using a social network (represented here by the Harvard friendship network introduced in Section 4). The social network is divided into communities using a state-of-the-art community detection algorithm, the Louvain method proposed by Blondel, Guillaume, Lambiotte, and Lefebvre, 2008, which maximizes a quality function known as modularity (Newman, 2006). Such a function evaluates the quality of a certain partition of the network into communities (in terms of number of communities and nodes per communities) under the assumption that communities are composed of nodes densely connected among one other and sparsely connected to the rest of the network. The expression of modularity is:

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(g_i, g_j) \tag{4}$$

where the communities are numbered starting from 1, and g_i is the number of the group to which node i belongs. The outcome of the employed community detection algorithm is a modularity value Q, a vector of integers reporting the assignment of nodes into communities $\mathbf{g}$, and a number of communities. Despite being widely used, one of the fundamental problems related to modularity maximization methods is the presence of degenerate or near-degenerate solutions (i.e., multiple high-quality solutions corresponding to different assignments of nodes to communities) that determine a certain level of ambiguity related to the task of community detection.

Let us suppose that the aforementioned service provider will provide an amount of services equal to the number of communities. Once the network has been divided into communities, nodes are assigned a certain metadata value, drawn from a uniform distribution, whereby the preference p_i of the node i for an offered service is represented. Each time that the preference of the node matches its community assignment (i.e., $p_i = g_i$) the service provider gets a return of 1. Therefore, the overall normalized return of the service provider can be written as:

$$R = \frac{\sum_i \delta(g_i, p_i)}{n} \tag{5}$$

The overall return associated to a certain partition depends on the correlation between the community assignment and the vector of preferences, as shown in Fig. 4. Since the assignment of preferences is drawn from a uniform distribution it doesn't show any particular correlation with the community assignment. Therefore, the actual assignment deriving from modularity maximization has no reason to be that

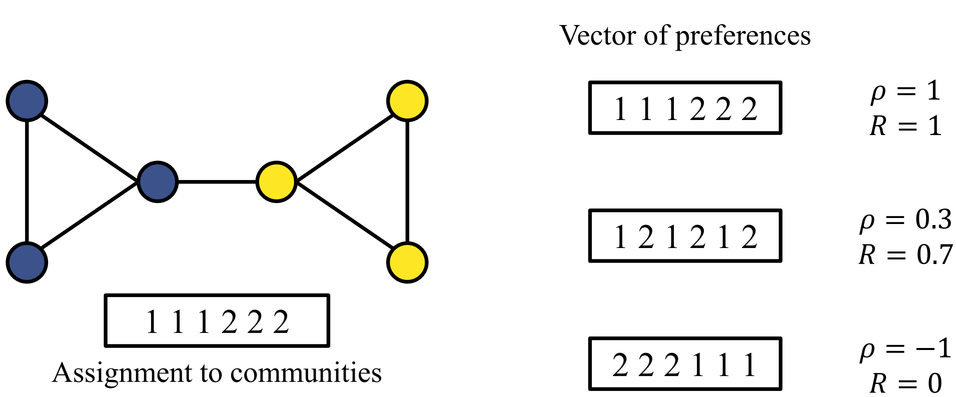

Fig. 4. Pictorial representation of community partition and preferences of the users. The example network is divided in two communities and the related assignment in reported in the vector beneath. The vectors on the right-hand side are examples users' vector of preferences and they are reported together with their correlation with the community assignment and associated return values. The correlation ρ is computed using the Pearson's correlation coefficient and the return R is computed using Eq. 5.

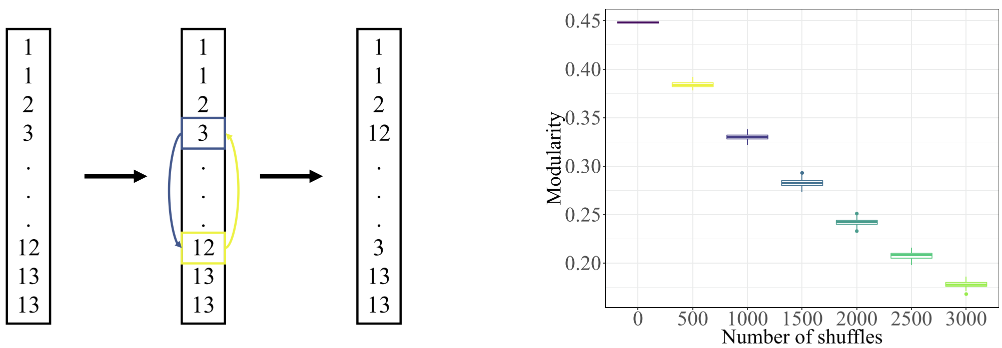

Fig. 5. Example of a shuffle of the community assignment and boxplots of modularity values associated to different shuffles. The left panel displays an example in which one element of the community assignment is shuffled. The right panel displays boxplots reporting the distribution of modularity for the community assignment obtained with the Louvain method. The assignment is shuffled 500, 1000, 1500, 2000, 2500, 3000 times for 250 instances. More than one instance of reshuffling is carried out so that an empirical distribution of the modularity Q and of the return R for each number of shuffles can be obtained.

guaranteeing the best return, while different assignments may provide better returns. Considering the vector of preferences of users fixed, it is possible to compute different values of R by exploring different partitions of the network in communities, as displayed in Fig. 4. Such an exploration problem is important in order to understand the multiple implications related to exploiting the network structure without being aware of its potential correlation with the node metadata.

Such an analysis provides a valid framework for testing hypotheses about users' preferences by means of a scenario analysis when we only have partial information about users' preferences (collected, for instance, via questionnaires) that we want to generalize to the whole network by testing different possible configurations. In summary, the network structure should be considered as the fundamental substrate to which other data should be posed to enrich it, with the scope of retrieving and exploiting the desired information.

The exploration of different communities assignments is conducted via a reshuffling of the assignment entries (a pictorial representation of the procedure is reported in Fig. 5 (*left*)) that entails, at each instance, the computation of a new modularity value Q and return R. The Harvard social network is divided into communities using the Louvain algorithm, which retrieves 13 communities and a value of modularity $Q\sim 0.48$. The elements of the related assignment are then shuffled 500, 1000, 1500, 2000, 2500, 3000 times, with 250 instances each for a total number of 1500 instances. The modularity values of the shuffles are reported in the boxplots of Fig. 5 (*right*). Each instance is associated with a certain return (reported in Fig. 6) and among the obtained results we are able to compute a Pareto front of the different (Q, R) couples. The Pareto front represents the set of the different optimal solution in terms of the Pareto optimality criterion, for which any of the solutions cannot be improved with respect to one objective without degrading the solution associated to another objective. The Pareto front turns out to be an important element to take into account when running community detection algorithms in order to consider more than one perspective and explore different viable options. The introduction of such information is useful in practical contexts, such as scenario analysis, and it may be helpful in addressing the ambiguity of the considered network outcome by avoiding decisions based on only one criterion. In Fig. 6 we observe that all the shuffled community assignments have a lower value of modularity than those deriving from the Louvain method; however, different returns are associated with such values. In fact, the assignment with the highest modularity is not associated with highest return; rather, it is conversely associated with an assignment with a low modularity score, as shown in Fig. 6.

The analysis of multiple scenarios underlines that the problem of finding communities can be embedded within a broader problem, still providing fundamental insights for the study of networks. The adoption of different criteria is helpful in order to address the ambiguity of community detection since it explains the extent to which the network structure is related to external information in terms of its elements, while also providing the possibility to analyze scenarios related to different decisions and strategies.

6. Conclusions

In this paper we have shown that network outcomes display a certain level of ambiguity related to their hardness of interpretation and to

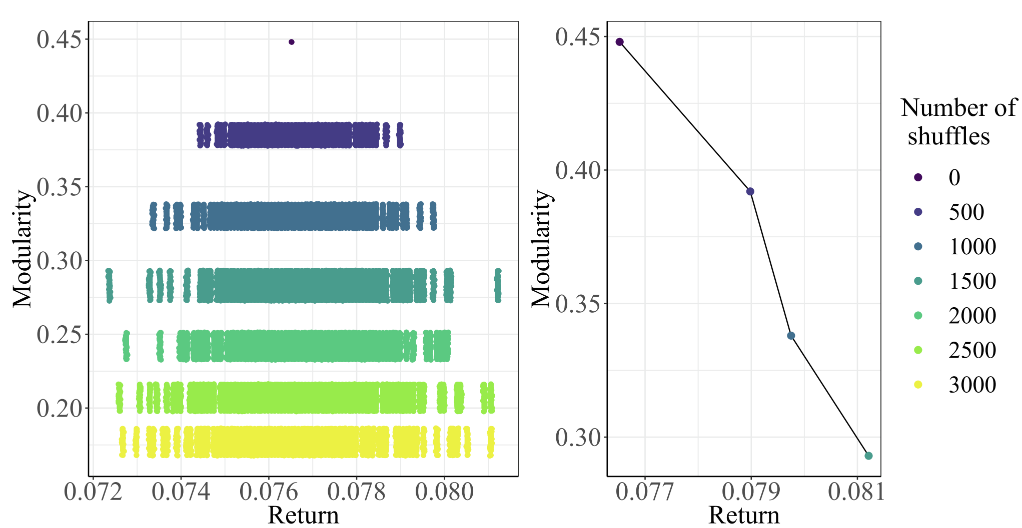

Fig. 6. Modularity values and returns of different shuffles and Pareto front associated to such values. The left panel reports a scatter plot of the values of modularity and return associated to different number of shuffles. The right panel reports the Pareto front deriving from the obtained shuffles.

the context in which we exploit such outcomes. We have also shown how a rationalization of the information content related to such measures, as well as the inclusion of different perspectives, can be helpful for addressing ambiguity. In more detail, we have proposed a novel method able to reduce the arbitrariness of ranking methods in the case of node-level outcomes, and have outlined a way to exploit the combination of structural and non-structural information to obtain improved measures in the case of network-level outcomes. The application of such methods to real data confirms the practical validity of the proposed approach. The framework provided in this paper can be extended to other outcomes (both at node level and at network level), and many different features, deriving from external data, can be included when networks are taken into account. The possibility to efficiently and rapidly address ambiguity of network outcomes is a key issue especially in business-related environments, where the opportunity of rapidly interpreting and exploiting the potential of network data can positively affect performances. Such an observation is in line with recent studies related to big data analytics for business that underline the potential of big data while pointing out certain restrictions due to the lack of technologies, tools and skills.

Other interesting concepts such as degeneracy, here used to define equivalent solutions of a problem realized by involving different nodes, may be considered. In fact, an interesting problem would be to understand the managerial and practical implications of degeneracy in terms of the possibility of reaching equivalent solutions involving different sets of nodes. An application where degenerate solutions could be exploited would be related to the task of team building where one could obtain equally valid groups made of different professionals.

Finally, more in-depth knowledge and the frequent use of complex networks in business environments would provide conspicuous benefits to companies and organizations. Such an effort would require the development of a wider amount of solution-oriented research at the junction between network science and business studies.

References

Abramo, G., D'Angelo, C. A., & Soldatenkova, A. (2017). An investigation on the skewness patterns and fractal nature of research productivity distributions at field and discipline level. *Journal of Informetrics, 11*(1), 324–335.

Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician, 27*(1), 17–21.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), P10008.

Burnes, B. (2005). Complexity theories and organizational change. *International Journal of Management Reviews, 7*(2), 73–90.

Cinelli, M. (2019). Generalized rich-club ordering in networks. *Journal of Complex Networks*.

Cinelli, M., Ferraro, G., & Iovanella, A. (2017). Structural bounds on the dyadic effect. *Journal of Complex Networks, 5*(5), 694–711.

Cinelli, M., Peel, L., Iovanella, A., Delvenne, & J. -C. (2019). *Network constraints on the mixing patterns of binary node metadata*. arXiv preprint arXiv:1908.04588.

Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review, 92*(4), 433.

Einhorn, H. J., & Hogarth, R. M. (1986). Decision making under ambiguity. *Journal of Business,* S225–S250.

Epstein, L. G., & Schneider, M. (2008). Ambiguity, information quality, and asset pricing. *The Journal of Finance, 63*(1), 197–228.

Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports, 659*, 1–44. community detection in networks: A user guide http://www.sciencedirect.com/science/article/pii/S0370157316302964.

Frisch, D., & Baron, J. (1988). Ambiguity and rationality. *Journal of Behavioral Decision Making, 1*(3), 149–157.

Good, B. H., De Montjoye, Y.-A., & Clauset, A. (2010). Performance of modularity maximization in practical contexts. *Physical Review E, 81*(4) 046106.

Håkansson, H., & Ford, D. (2002). How should companies interact in business networks? *Journal of Business Research, 55*(2), 133–139.

Kahn, R. L., Wolfe, D. M., Quinn, R. P., Snoek, J. D., & Rosenthal, R. A. (1964). *Organizational stress: Studies in role conflict and ambiguity*.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*.

Lü, L., Medo, M., Yeung, C. H., Zhang, Y.-C., Zhang, Z.-K., & Zhou, T. (2012). Recommender systems. *Physics Reports, 519*(1), 1–49.

Newman, M. E. (2003). Mixing patterns in networks. *Physical Review E, 67*(2) 026126.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences, 103*(23), 8577–8582.

Newman, M. E. (2010). *Networks: An Introduction.* New York: Oxford University Press.

Noordegraaf, M., & Abma, T. (2003). Management by measurement? public management practices amidst ambiguity. *Public Administration, 81*(4), 853–871.

Peel, L., Delvenne, J.-C., & Lambiotte, R. (2018). Multiscale mixing patterns in networks. *Proceedings of the National Academy of Sciences, 115*(16), 4057–4062.

Peel, L., Larremore, D. B., & Clauset, A. (2017). The ground truth about metadata and community detection in networks. *Science Advances, 3*(5) e1602548.

Pfeffer, J. (1977). The ambiguity of leadership. *Academy of Management Review, 2*(1), 104–112.

Piraveenan, M., Prokopenko, M., & Zomaya, A. (2008). Local assortativeness in scale-free networks. *EPL (Europhysics Letters), 84*(2), 28002.

Piraveenan, M., Prokopenko, M., & Zomaya, A. (2010). Local assortativeness in scale-free networks. *EPL (Europhysics Letters), 89*(4), 49901.

Powell, T. C., Lovallo, D., & Caringal, C. (2006). Causal ambiguity, management perception, and firm performance. *Academy of Management Review, 31*(1), 175–196.

Schubert, A., Glänzel, W., & Braun, T. (1987). Subject field characteristic citation scores and scales for assessing research performance. *Scientometrics, 12*(5–6), 267–291.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research, 70*, 263–286.

Traud, A. L., Mucha, P. J., & Porter, M. A. (2012). Social structure of facebook networks. *Physica A: Statistical Mechanics and its Applications, 391*(16), 4165–4180.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*, *Vol*, Cambridge University Press8.

Zhou, S., & Mondragón, R. J. (2004). The rich-club phenomenon in the internet topology. *IEEE Communications Letters, 8*(3), 180–182.