Sharpening the definition of centrality

Matthieu Roy
LAAS-CNRS, Toulouse

Gilles Tredan
Deutsche Telekom/TU-Berlin

1 Introduction

Since Web 2.0, online content has become social. The popularization of Internet access added to the ease of producing online content have brought social communities to go online.

This penetration of social assets in the modern internet has a wide impact, which naturally led computer scientists to analyze these online social networks. The promise of this research is for instance, from a network point of view, to discover *who* will exchange *which* information, and *when*. These three pieces of information are obviously central in network design.

Among the available tools to dig and leverage information out of complex interaction structures, probably the most interesting one is the *centrality* notion. Originally defined to capture "the importance of an actor in a network"[4], it has since found many application in various fields, such as management, physics, or biology. In fact, centrality appears to be a much more fine-grain tool to understand the inner structure of social networks than other information such as degree distribution. Recently, Newmann even used it to study one of the core elements of online social networks: communities [9]

However, the apparently simple notion of centrality hides many different concepts. Despite its success, centrality lacks a proper definition, and may more be referred to as an abstract family of measures rather than to a single notion. There is unfortunately no generic definition of a centrality as a set of mathematical properties, as there is no absolute definition of the "importance" of an actor. So for each subjective definition of importance, an ad-hoc centrality definition has been provided.

In this paper, we prospect for a more precise definition of centrality. We believe that the mathematical usefulness of this tool could greatly benefit from such definition.

2 Related works

Network science has recently mutated to a multidisciplinary science field. Indeed, this discipline spans over multiple domains of science: biology, sociology, physics, economy, computer science and mathematics. It seems that most of science domains, beyond understanding the isolated elements that define their domain, have at some point to understand the interactions that occur between these elements contextualized together in a system.

Centralities' past reflects these multiples influences. The first traces of works quantifying the role of a node with respect the network it belongs originate from Camille Jordan (1869), a mathematician whose objective was to detect the "center of a graph". However, the first use of centrality as we now know it is attributed to Bavelas and Leavitt in the late 40's, as a tool to study organisations and assess productivity. This first use was largely followed in a wide spectrum of scientific fields, such as genetics, biology, physics, sociology. In [5], Freeman relates the centrality's trajectory among these diverse fields.

Despite its success, centrality remains poorly defined. Borgatti and Everett [1] explain that "the only thing people agree about a centrality is that it is a node-level measure". Sabidussi [10] issued a first attempt to define formal criteria that a centrality has to match. However, as Borgatti reports, these criteria exclude several known centralities, such as the betweenness centrality.

Most of other graph-theoretic approaches to centrality rather try to define categories or to group centralities into various classes, depending on the nature of the graph property they capture. Incorporating Freemans' early 3 core concepts of centrality [4], Borgatti and Everett [1] describes centralities along 4 dimensions. These insightful works thus provide an atlas of current centralities, but fail to refine the original and simplistic centrality definition.

Finally, it is interesting to notice diverse variations of "core centralities" (as defined by Freeman). For examples, Brandes [2] explores several variations of the betweenness, Dolev *et al.* [3] express centralities as routings, and Kermarrec *et al.* [7] define a new centrality, the second order centrality, that aims to be exploited in a distributed setting. All these recent works suggest that centralities are still a tool under heavy development, and that known centralities atlas may become quickly outdated.

3 Some questions about centrality

We explore here some directions that may help to capture the definition of centrality. With some rough sketches, we try to foster reflection on the centrality and its use. As a starting point, we consider that a centrality is a node-level measure. Thus, for each graph, it produces a vector of centralities. Let $\mathcal{G}$ be the space of binary graphs[1].Then a centrality c is an application that, given a graph $G \in \mathcal{G}$, returns a measure on its vertices $V(G)$:

$$\begin{aligned} c : \mathcal{G} &\rightarrow {\mathbb{R}^+}^{V(\mathcal{G})} \\ G &\mapsto {\mathbb{R}^+}^{V(G)} \end{aligned}$$

Centrality and automorphisms : Indeed, concerning the mathematical properties of centralies, a first thing to notice is that centrality is robust to automorphism, i.e., it is invariant by permutation of nodes identities (this is the classical definition of anonymity in distributed systems).

More formally, consider two graphs G_1 and G_2 such that there exists an automorphism that transforms G_1 in G_2 (i.e., G_1 is symmetric, and G_2 is isomorphic to G_1). Then, we believe that for any centrality c, $c(G_1) = c(G_2)$, since the structure remains unchanged. Thus, c is not injective. This is a first refinement of the original definition.

It is maybe possible to use automorphisms to define centralities more precisely. Consider for example a line graph of n nodes, where each node i, $0 < i < n-1$ is linked to $i+1$ and $i-1$. It is to us hard to imagine a centrality c such that, given a node i , $(i \neq \frac{n}{2})$, we have $c(i) \notin [c(i-1), c(i+1)]$. Is this a criterion for a centrality ?

Is there a surjective centrality ? Since we show that a centrality c is not injective, a natural question is to ask whether c can be surjective. In other words, can we always found a graph that generates a given centrality distribution ? Since centralities usually map a finite (or at least countable) set of vertices to a dense compact space (such as an interval of $\mathbb{R}^+$), obviously a centrality cannot be surjective.

We can thus try to restrict c's codomain. This work was done for degree centrality, since the problem of characterising degree distributions that define a graph (*graphical* distributions) was solved in 1960 by Erdös and Gallai (the interested reader can see *e.g.* [6]). However, finding the

[1]valued graphs often require centralities to be redefined. This redefinition is not always straightforward.

codomains where the other known centralities (such as betweenness) are surjective remains an open question.

Can centralities define graphs? The opposite question, i.e., defining a graph based on information on centralities, seems difficult to solve in the general case: centralities, by nature, provide reduced information on the inner structure of a given network.

Can a graph be uniquely defined by a certain amount of centrality distributions ? Although intuitively the answer provided would be negative, let us consider a 3 nodes graph with degree centralities $(1, 1, 2)$. It is easy to deduce from this information that it defines a unique graph (to an isomorphism): a "line" of 3 nodes. Now consider a larger graph with more than 3 nodes, but add other centralities (*e.g.* betweenness or eigenvector). This approach is very close to [8], where the authors claim to be able to reconstruct a graph using degree correlations.

An open question is to determine the minimal amount of information on centralities *needed* to uniquely define a graph.

References

[1] S.P. Borgatti and M.G. Everett. A graph-theoretic perspective on centrality. *Social Networks*, 28(4):466–484, 2006.

[2] U. Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2):136–145, 2008.

[3] S. Dolev, Y. Elovici, and R. Puzis. Routing Betweenness Centrality. *Technical Report 2009-09*, 2009.

[4] L. C. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1:215–239, 1978.

[5] L. C. Freeman. Going the wrong way on a one-way street: Centrality in physics and biology. *Journal of Social Structure*, 2008.

[6] SL Hakimi. On realizability of a set of integers as degrees of the vertices of a linear graph II. Uniqueness. *Journal of the Society for Industrial and Applied Mathematics*, 11(1):135–147, 1963.

[7] A.M. Kermarrec, E. Le Merrer, B. Sericola, and G. Trédan. Evaluating the Quality of a Network Topology through Random Walks. *Distributed Computing*, pages 509–511.

[8] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat. Systematic topology analysis and generation using degree correlations. In *Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, page 146. ACM, 2006.

[9] Mark Newman. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA*, 103:8577–8582, 2006.

[10] G. Sabidussi. The centrality index of a graph. *Psychometrika*, 31(4):581–603, 1966.