

Spatial networks and small worlds

Luca Lombardo

Dicember 2022

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Contents

1	Introduction	2
2	Theoretical background on centrality measures	4

1 Introduction

Given a social network, which of its nodes are more central? This question has been asked many times in sociology, psychology and computer science, and a whole plethora of centrality measures (a.k.a. centrality indices, or rankings) were proposed to account for the importance of the nodes of a network.

These networks, typically generated directly or indirectly by human activity and interaction (and therefore hereafter dubbed social”), appear in a large variety of contexts and often exhibit a surprisingly similar structure. One of the most important notions that researchers have been trying to capture in such networks is “node centrality”: ideally, every node (often representing an individual) has some degree of influence or importance within the social domain under consideration, and one expects such importance to surface in the structure of the social network; centrality is a quantitative measure that aims at revealing the importance of a node.

Among the types of centrality that have been considered in the literature, many have to do with distances between nodes. Take, for instance, a node in an undirected connected network: if the sum of distances to all other nodes is large, the node under consideration is peripheral; this is the starting point to define Bavelas’s closeness centrality [1], which is the reciprocal of peripherality (i.e., the reciprocal of the sum of distances to all other nodes).

The role played by shortest paths is justified by one of the most well-known features of complex networks, the so-called small-world phenomenon. A small-world network [2] is a graph where the average distance between nodes is logarithmic in the size of the network, whereas the clustering coefficient is larger (that is, neighborhoods tend to be denser) than in a random Erdős-Rényi graph with the same size and average distance. The fact that social networks (whether electronically mediated or not) exhibit the small-world property is known at least since Milgram’s famous experiment [] and is arguably the most popular of all features of complex networks. For instance, the average distance of the Facebook graph was recently established to be just 4.74 [3].

Definitions and conventions

From now on, we consider directed graphs defined by a set N of n nodes and $A \subseteq N \times N$ of arcs. We write $x \rightarrow y$ when $(x, y) \in A$ and call x and y the source and the target of the arc, respectively.

Aim of the project

The Aim of the project is to study the small-world phenomenon in location-based (social) networks. As test cases, we consider three real-world datasets: Brightkite, Gowalla and Foursquare. In the next sections, we will describe the datasets and the methodology we used to extract the networks from them.

We are interest in analyzing 4 different centrality measures:

- Distribution of Degree
- Clustering coefficient
- Average Path Length
- Betweenness Centrality

2 Theoretical background on centrality measures

Centrality is a fundamental tool in the study of social networks: the first efforts to define formally centrality indices were put forth in the late 1940s by the Group Networks Laboratory at MIT directed by Alex Bavelas [1]; those pioneering experiments concluded that centrality was related to group efficiency in problem-solving, and agreed with the subjects' perception of leadership. In the following decades, various measures of centrality were employed in a multitude of contexts.

Geometric measures

We call geometric those measures assuming that importance is a function of distances; more precisely, a geometric centrality depends only on how many nodes exist at every distance. These are some of the oldest measures defined in the literature.

In-degree centrality In-degree, the number of incoming arcs $d^-(x)$, can be considered a geometric measure: it is simply the number of nodes at distance one¹. It is probably the oldest measure of importance ever used, as it is equivalent to majority voting in elections (where $x \rightarrow y$ if x voted for y). In-degree has a number of obvious shortcomings (e.g., it is easy to spam), but it is a good baseline.

Other notable geometric measures that we will not explore in this project, are *closeness centrality*, (which is the reciprocal of the sum of distances to all other nodes, and *betweenness centrality*, which is the number of shortest paths that pass through a node), *Lin's index* (which is the sum of the distances to all other nodes), and *Harmonic Centrality* (which is a generalization of the closeness centrality).

¹Most centrality measures proposed in the literature were actually described only for undirected, connected graphs. Since the study of web graphs and online social networks has posed the problem of extending centrality concepts to networks that are directed, and possibly not strongly connected, in the rest of this paper we consider measures depending on the incoming arcs of a node (e.g., incoming paths, left dominant eigenvectors, distances from all nodes to a fixed node). If necessary, these measures can be called “negative”, as opposed to the “positive” versions obtained by considering outgoing paths, or (equivalently) by transposing the graph.

References

- [1] Alex Bavelas. A mathematical model for group structures. *Applied Anthropology*, 7(3):16–30, 1948.
- [2] Reuven Cohen and Shlomo Havlin. *Complex Networks: Structure, Robustness and Function*. Cambridge University Press, 2010.
- [3] Stanley Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.
- [4] Dingqi Yang, Daqing Zhang, Vincent. W. Zheng, and Zhiyong Yu. Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(1):129–142, 2015.
- [5] Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna. Four degrees of separation, 2011.