Physica D 267 (2014) 58–67

Contents lists available at ScienceDirect

Physica D

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

Composite centrality: A natural scale for complex evolving networks

A.C. Joseph *, G. Chen [1]

City University of Hong Kong, Department of Electronic Engineering, Centre for Chaos and Complex Networks, 83 Tat Chee Avenue, Kowloon, Hong Kong Special Administrative Region

HIGHLIGHTS

- A procedure for statistical measure standardisation is proposed.
- A novel framework of composite measures for complex evolving networks is proposed.
- Demonstration of its working by investigating real data from the WTW and the WMW.
- The composite measure is shown to follow a parameter-free std. normal distribution.
- Simulations demonstrate applicability of the concept to large and general data sets.

ARTICLE INFO

Article history:
Available online 22 August 2013

Keywords:
Weighted directed network
Evolving network
Complex system
Unified scale
World trade web
World migration web

ABSTRACT

We derive a composite centrality measure for general weighted and directed complex networks, based on measure standardisation and invariant statistical inheritance schemes. Different schemes generate different intermediate abstract measures providing additional information, while the composite centrality measure tends to the standard normal distribution. This offers a unified scale to measure node and edge centralities for complex evolving networks under a uniform framework. Considering two real-world cases of the world trade web and the world migration web, both during a time span of 40 years, we propose a standard set-up to demonstrate its remarkable normative power and accuracy. We illustrate the applicability of the proposed framework for large and arbitrary complex systems, as well as its limitations, through extensive numerical simulations.

1. Introduction

Starting with the work of Watts, Strogatz, Barabási, Albert and Newman [1–3], the investigation of complex networks has attracted an inflationary amount of attention from numerous research fields due to their ubiquity in the real world [4–6]. One of the fascinations lies in some elegant and efficient descriptions of very different complex systems under the general framework of modern graph theory [7,8] pioneered by Erdös [9].

As the awareness of several common dynamic characters of many real-world networks rises, ever more effort has been devoted to understanding the temporal evolution of complex networks. Prominent examples of evolving networks are the internet and social networks [5], transportation networks [10], the world trade web [11,12] and most recently climate networks [13]. Any of such networks is generated by and/or hosts an underlying flow between its nodes, such as information, contacts, goods or diseases. Considering the description and analysis of evolving network structures, most efforts have been made regarding network modelling [14–18], while the development of sophisticated analysis tools and methodologies has seen less progress. This yet will be the topic of this work.

The characterisation and classification of general complex systems and especially complex evolving networks pose three major challenges:

- *Uniformity*: There is a large variety of network measures stretching over wide numerical ranges, but there is no standardised procedure today to consistently consider several measures simultaneously.
- *Variability*: Observed over time, many complex networks show growth (change of the number of nodes) and evolution (change of the topology). Network measures often depend explicitly on these quantities, which complicates a coherent temporal analysis.
- *Comparability*: There is no unified scale on which one can compare results originating from different networks.

* Corresponding author.
E-mail addresses: andreas.ch.joseph@gmail.com (A.C. Joseph), eegchen@cityu.edu.hk (G. Chen).
[1] Tel.: +852 3442 7922; fax: +852 3442 0562.

0167-2789/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physd.2013.08.005

Furthermore, [19,20] classified the underlying flows and the corresponding graph measures in terms of their physical or graph-theoretic properties, respectively. In a given situation, the exact details of the underlying flows might not be well-understood, and a multi-dimensional analysis of graph measures allowing for simultaneous evaluations is desirable.

Motivated by this and the above-stated general problems, we propose here a new centrality framework, called *composite centrality* (CC). Generally, the notion of centrality can be understood as a measure quantifying the participation of a node (or any other component) in the underlying flow structure of a network [19]. This will also be the point of view we adopt in this work. The idea behind the CC-framework is that one first defines a set of characteristics of interest, and then chooses appropriate network (centrality) measures. The major complications when considering multiple network measures are different (often arbitrary) numerical scales and variously shaped distributions such as distributions with and without heavy tails. We therefore implement a *standardisation procedure* involving a non-linear transformation and statistical normalisation. Relying on statistical methods, uniformity and variability are accounted for. Standardised measures can be combined using *invariant inheritance schemes* to form new standardised measures carrying abstract physical meanings, which we call *composite centrality*. It turns out that final CC-scores for different set-ups and different networks are well-approximated by the standard normal distribution with a zero mean and a unit variance. This is what we call a universal scale to compare scores for different set-ups and even different networks across time (comparability).

This paper is structured as follows. In Section 2, we give a short introduction to the relevant terminology from graph theory and propose a recipe for (graph) measure standardisation. In Section 3, we present the CC-framework and introduce a specific standard framework. We demonstrate the working of the proposed set-up by considering two cases of the world trade web and the world migration web, both during a time span of 40 years. Furthermore, a graphical tool, which we call the *network genetic fingerprint*, is introduced. It allows for efficient analysis and monitoring of composite centrality scores. In Section 4, we discuss the validity and limitations of the proposed framework through large-scale simulations. We finally conclude the study in Section 5.

2. Preliminaries

2.1. Graph theory

In this section, we give a short introduction to the parts of graph theory that are needed in the following. Explicit formulas will be given only if deemed necessary. For a more detailed introduction, we refer to [4–7].

Graph theory provides a general mathematical framework to represent and quantify complex networks and their properties. A weighted and directed network can be represented by a graph $G = (V, E)$, where $V = \{v_1, \ldots, v_N\}$ is the set of $N \geq 2$ nodes (vertices) in the graph. $E\left(w_{ij} > 0 \mid i, j \in \{1, \ldots, N\}\right)$ is the set of weighted edges from node v_i to node v_j, with $N_e = \mathrm{ord}(E)$ denoting the number of edges irrespective of their weights. The whole graph can be represented by a real weight matrix, $W = [w_{ij}] \in \mathbb{R}^{N\times N}$ ($w_{ij} \neq w_{ji}$, in general). We do not allow for self-loops here, i.e. $w_{ii} = 0$ for all $i \in \{1, \ldots, N\}$. The ith row or column represents the out- or in-strength distribution of node i, respectively. The *total strength* of a node i, denoted by s_i, is the sum of in-strengths s_i^{in} and out-strengths s_i^{out} of that node. It represents a generalisation of the degree centrality (number of adjacent edges) in an undirected and unweighted graph. The degree of a node can be obtained from the underlying simple (non-weighted, non-directed, no self-loops) graph through its adjacency matrix $A = [a_{ij}] \equiv [a_{ji}] \in \{0, 1\}$, where $a_{ij} = 1$ if there is an edge between node i and node j, but $a_{ij} = 0$ otherwise. Likewise, since self-loops are not allowed, one has $a_{ii} = 0$ for all $i \in \{1, \ldots, N\}$. The degree of node i is given by the ith row or column sum of A. Strength and degree of a node can be interpreted as two measures for local connectivity either considering weighted or unweighted graphs, respectively. Here and later, we refer to measures over weighted networks as being of *quantitative* nature and measures over unweighted networks as being *qualitative*. The difference between both levels of complexity is summarised under the notion of (edge) *texture*. It is said that there is a connection between any two nodes i and j in G if there exists a directed path p_{ij} from i to j ($p_{ij} \neq p_{ji}$, in general). A directed graph is said to be *strongly-connected* if there exists a directed path between any two nodes. This means that the weight matrix W and the adjacency matrix A are both irreducible. A measure for the connectivity of a graph on the global scale is the *edge density* $\rho_e = N_e/\left(N^2 - N\right)$ (number of actual edges divided by number of possible edges), while on a local scale the embedding of a node can be expressed via its *clustering coefficient*. On the adjacency level, the clustering coefficient of a node is defined as the number of actual connections within its neighbours over the number of possible connections among them. A further important measure to quantify the participation of a node i into the path-structure of a network indicating overall connectivity is the *average shortest path length* per other node $l_i = \langle l_{ij}\rangle_i$ (or *farness*), i.e. the average number of steps (over unweighted edges) which it takes to get from node i to any other node j. A generalisation to weighted edges is straightforward, once one relates edge weight to distance. Note that in a directed graph the shortest path between two nodes is generally not symmetric, i.e. $l_{ij} \neq l_{ji}$. The *diameter* ø of a graph is defined as the maximal shortest path between any two nodes. The *maximal flow* f_{ij} between two nodes i and j is the maximal capacity that can be transported parallelly from node i to node j via the whole graph for networks in which the edge weight can be interpreted as representing some form of capacity [4], e.g. bandwidth for electronic data transmission (again $f_{ij} \neq f_{ji}$, in general).

Graph asymmetry is a measure for the difference between w_{ij} and w_{ji} on a global scale, i.e. for the overall weight balance. We define it as

$$A_D = \frac{\|W - W^T\|_F}{2\,\|W\|_F} \in [0, 1], \tag{1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. The *algebraic connectivity* λ_1 of the normalised Laplacian $L_N = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}$, where $D = \mathrm{diag}(s_1, \ldots, s_N)$ is a diagonal matrix consisting of the nodes' strengths, is the smallest non-zero eigenvalue, while the Laplacian always has a single zero eigenvalue for the case of only one single connected component. It is a measure for the robustness of a graph against node removal (failure) [4,21]. A further measure is *assortativity* [4,5], $As \in [-1, 1]$, which describes the overall homogeneity of connections indicating if weak/strong nodes are preferentially coupled to other weak/strong nodes (or vice versa), resulting in a positive (or negative) As value. *Eigenvector centrality* of a node i is defined as the ith entry of the eigenvector corresponding to the largest eigenvalue of the underlying graph's adjacency matrix. It measures how well a node is connected to the whole graph or to other well-connected (high-scoring) nodes recursively [4].

2.2. Measure standardisation

Different node measures (e.g. centrality measures) generally span wide and different numerical ranges, show different levels of variation and exhibit variously shaped distributions, which makes them difficult to compare.[2] For instance, it is difficult to

[2] Here and throughout, we only consider node measures. A generalisation to edge measures is straightforward.