Angela Bohn and Norbert Walchhofer and Patrick Mair and Kurt Hornik
Social Network Analysis of Weighted Telecommunications Graphs

Original Citation:

This version is available at: http://epub.wu.ac.at/708/
Available in ePub°WU°: March 2009

ePub°WU°, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.
Social Network Analysis of Weighted Telecommunications Graphs

Angela Bohn, Norbert Walchhofer, Patrick Mair, Kurt Hornik

Department of Statistics and Mathematics
Wirtschaftsuniversität Wien

Research Report Series
Report 84
March 2009

http://statmath.wu-wien.ac.at/
Social Network Analysis of Weighted Telecommunications Graphs

Angela Bohn Norbert Walchhofer Patrick Mair
Kurt Hornik

Abstract

SNA provides a wide range of tools that allow examination of telecommunications graphs. Those graphs contain vertices representing cell phone users and lines standing for established connections. Many sna tools do not incorporate the intensity of interaction. This may lead to wrong conclusions because the difference between best friends and random contacts can be defined by the accumulated duration of talks. To solve this problem, we propose a closeness centrality measure (ewc) that incorporates line values and compare it to Freeman’s closeness. Small exemplary networks will demonstrate the characteristics of the weighted closeness compared to other centrality measures. Finally, the ewc will be tested on a real-world telecommunications graph provided by a large Austrian mobile service provider and the advantages of the ewc will be discussed.

1 Introduction

So far, the so called Freeman’s closeness centrality measure could only be calculated for unweighted graphs. As in many applications the line values are important for the understanding of networks, we introduce an edge-weighted closeness (ewc). Contrary to other versions of weighted closeness that can be found in literature, the ewc incorporates not only line values but also distances. The ewc code (Appendix A) for R [R Development Core Team, 2008] will be provided. The ewc will be compared to Freeman’s closeness and other centrality measures. Furthermore, it will be tested for a real-world cellphone calls graph. We will show that the ewc detects nodes that have very close contacts to their friends. On this note, the ewc is a counterpart of betweenness centrality which identifies people with rather loose connections to many separated groups.

2 Literature Review

2.1 History of Closeness Centrality

A centrality measure based on distances was first introduced by Bavelas [1950]. He noticed that in different kinds of networks containing the same number of vertices, the sum of distances differs. In addition, he showed on the basis of examples that the calculation of the absolute distance between a vertex and all other vertices does not suffice to describe a its role in the network. Thus he suggested to use the proportion of the sum of all distances in the network.
\[ \sum d(x, y) \] and the sum of the distances from a certain vertex \( i \) to all others \( j \) to describe \( i \)'s importance: \[ \frac{\sum d(x, y)}{\sum d(i, j)} \]. Beauchamp (1965) had the idea of rescaling Bavelas’ closeness to the 0-1 interval. His “relative closeness” \( RC(i) \) of vertex \( i \) is

\[ RC(i) = \frac{n - 1}{\sum_j d(i, j)}, \]  

where \( n \) is the network size (=number of vertices). This is the formula commonly used for closeness centrality, also called Freeman’s closeness. Moxley and Moxley (1974) focused on the problem of disconnected graphs. While Flament (1963) defined the distance between two vertices that cannot reach each other to be infinite and Beauchamp (1965) suggested to leave them undefined, Moxley and Moxley (1974) assigned a penalty to each vertex that is not able to reach every other vertex. This penalty has to be larger than the maximum distance (=number of vertices-1). Freeman (1979) summarized and structured the centrality measures existing so far like degree, betweenness and closeness. Therefore, he is the most cited author when it comes to closeness centrality. He also introduced a measure of (closeness) centralization, a measure of the extend to which a network has a center and a periphery.

### 2.2 Freeman’s Closeness Centrality

Although Beauchamp (1965) created the closeness used today, it is usually referred to as Freeman’s closeness. The formula for closeness is

\[ C_C(i) = \frac{n - 1}{\sum_j d(i, j)} \]  

where \( i \) is the observed vertex, \( n \) is the network size (=number of vertices), \( j \) is a vertex that is connected to \( i \) by a path and \( d(i, j) \) is the distance between \( i \) and \( j \) (=length of the shortest path between \( i \) and \( j \)).

The closeness is a point-centrality measure: it is used to measure the importance of a network member. It is based on the idea that, the farther a vertex \( j \) is away from the observed vertex \( i \), the less it should contribute to \( i \)'s importance. Thus, the closeness of \( i \) is large if the sum of distances from \( i \) to all other nodes is small. If \( i \) is located at the periphery of the network, the sum of distances is large and the closeness is small.

If we had a network with the adjacency matrix

\[ \text{network1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 10 \\ 0 & 10 & 0 \end{pmatrix}, \]

the two outer vertices had the same closeness, because only distances are taken into account and weights are ignored.

![Figure 1: Example of Freeman’s Closeness](image-url)
2.3 Past Research Concerning “Weighted Closeness”

So far, a “weighted closeness” has been used in seven papers presenting five different versions of it. Although all of them start with Freeman’s closeness (Freeman, 1979), each version of weighted closeness comes out different because it was tailored to the papers’ purposes:

Toivonen et al. (2006) analyzed the relationships between trustors and trustees from a network perspective. They use a weighted closeness as a measure of closeness between two people, not as a measure for a single person.

Cornwell (2005) suggested a different method to make Freeman’s closeness applicable to disconnected graphs. He called this new closeness “complement-weighted closeness”, because he uses the complement of a graph in order to calculate it. The complement $G_C$ of a graph $G$ is a network consisting of the same number of nodes as $G$ and of those lines that do not exist in $G$. He can use the complement $G_C$ to calculate the closenesses for a disconnected graph $G$, because in an undirected or symmetric network, the complement $G_C$ is always a connected graph.

Fowler (2005, 2006a,b) analyzed the co-sponsorship network of the US legislative from 1973 to 2004. He used the formula for the normal unweighted closeness and defined a weight $w(i, j)$ indicating the intensity of the relationship between a sponsor $j$ and a co-sponsor $i$:

$$w(i, j) = \frac{\sum_{l} a(i, j, l)}{c(l)},$$

where $l$ is a co-sponsored bill, $a(i, j, l)$ is a binary indicator taking the value 1 if legislator $i$ co-sponsors bill $l$ that is sponsored by legislator $j$ and 0 otherwise and $c(l)$ is the number of co-sponsors of bill $l$. In a next step, Fowler set the distance $d(i, j)$ as $\frac{1}{w(i, j)}$. Thus, he defined the weights as distances and this way, he could use Freeman’s closeness as a “weighted closeness”.

Erath et al. (2007) used a weighted closeness measure in order to find important points in the Swiss transport network. They used the importance of a destination as weight, such that the weight is attributed to a vertex and not to a line. They defined the closeness $C_C(i)$ of a destination $i$ as

$$C_C(i) = \sum_{j \neq i} \frac{\sum_{j \neq i} W_j TT_{ij}}{\sum_{j \neq i} W_j},$$

where $W_j$ is the importance of destination $j$ and $TT_{ij}$ is the travel time between $i$ and $j$.

Newman (2001) analyzed scientific collaboration networks. He calculated “the weighted version of the closeness centrality measure [...] i.e., the average weighted distance from a vertex to all others”. He did not state any formula, but in the course of personal communication, he said that he used a distance measure which was based on edge lengths. The edge lengths were the reciprocals of their weights. Closeness was then the average of the lengths of these paths over all vertices. From this description we assume that his formula was

$$C_{NC}(i) = \frac{\sum_{j} \frac{1}{d(i, j)}}{d(i, j)},$$
where \( w(i, j) \) is the weight of the edges needed to go from \( i \) to \( j \). In this formula, higher line values lead to smaller closeness. So, if we calculate Newman’s weighted closeness for \textit{network1} we get:

\[ C_{RC}(i) = \frac{n - 1}{\sum_j w(i, j)}. \]  

Consequently, high line values reduce the closeness, which is an undesired effect. Additionally, as the distances disappear from the formula, each vertex \( j \)’s contribution to \( i \)’s closeness depends on the sum of line values on the path from \( i \) to \( j \), no matter how far \( j \) is away. Figure 3 shows the weighted closeness of the \textit{sna} Package calculated for \textit{network1}:

\[ C_{ewc}(i) = \frac{\sum_j llv(i, j) d(i, j)}{\text{max}(lv)(n - 1)}, \]  

where \( \text{max}(lv) \) is the maximum of the line values occurring in the observed network, \( n \) is the network size, \( j \) are all vertices that can be reached by \( i \), \( llv(i, j) \) is the average\(^1\) last line value occurring on the path from \( i \) to \( j \) and \( d(i, j) \) is

\(^1\)There can be several shortest paths. In this case we take the average of the last line values.
the distance between $i$ and $j$. Read Wasserman and Faust (1997) to see how to obtain the distance matrix.

The denominator $\max(lv)(n - 1)$ standardizes the values, so an ewc close to 1 is associated with a high centrality, whereas an ewc close to 0 indicates low centrality.

If we calculate the ewc for network1, we get:

$$
\begin{array}{c}
0.3 \\
0.55 \\
0.525 \\
\end{array}
$$

Figure 4: Example of Edge-Weighted Closeness

Now the right vertex has a higher centrality because it has a larger line value than the left node.

Line values and degree contribute in equal measure to the ewc. One degree weights the same as a line value of one. (This can be changed by transforming the line values.) Example:

<table>
<thead>
<tr>
<th></th>
<th>network2 = $\begin{pmatrix} 0 &amp; 2 &amp; 0 \ 2 &amp; 0 &amp; 3 \ 0 &amp; 3 &amp; 0 \end{pmatrix}$</th>
<th>network3 = $\begin{pmatrix} 0 &amp; 5 \ 5 &amp; 0 \end{pmatrix}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before standardization:</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>After standardization:</td>
<td>0.83</td>
<td>1</td>
</tr>
</tbody>
</table>

Tabular 1: ewc for different line values and degrees in comparison

A vertex with two adjacency lines with values two and three has the same non-standardized ewc as a vertex with one adjacency line with value five. However, the standardization causes the ewc values to be not the same because $\max(lv)$ and $n$ change.

The code for R is provided in Appendix A.

4 EWC Compared to Other Centrality Measures

The graphs show the mailing list conversations of R-devel\textsuperscript{2} in January 2008, visualizing six different centrality measures. In each plot, the node having the largest centrality is marked with 1, the vertex with the second highest centrality is labeled with 2, and so on. The vertex size corresponds to the actual centrality. Broad dark lines indicate high communication intensity while narrow light lines

\textsuperscript{2}R-devel is a mailing list for R developers. [http://stat.ethz.ch/mailman/listinfo/r-devel](http://stat.ethz.ch/mailman/listinfo/r-devel) The network can be downloaded from [http://www.angela-bohn.de/data.html](http://www.angela-bohn.de/data.html)
Figure 5: EWC compared to five other centrality measures

The blue graph in the top left-hand corner shows the ranked results of ewc. The large vertex at the bottom right corner has the highest ewc. Therefore so it is marked with 1. The node labeled with 2 has the second highest ewc because mark weak ties.
the large number of direct neighbors contributes to its ewc. The vertex with the third highest ewc has a slightly smaller line values than Node 1 and a smaller degree than Node 2. If we look at the results of Freeman’s closeness by comparison, we see that the large vertex in the center is the number 1 vertex. This is due to the fact that line values are not taken into account. The vertex with the highest ewc has only the 9th highest closeness. In contrast, the second highest closeness node is not very important according to the 3, where it is on the fifth position.

If we compare the ewc to the weighted closeness of the R sna package, we see that the latter does not yield the desired results. While the fact that the node in the center is number 1 can be discussed, the vertex on the bottom right corner having high line values should not be on the 17th position.

The main difference between weighted degree and ewc is the fact that the degree considers direct neighbors only. Thus, the results are similar for networks with small maximal distance like this one.

The betweenness looks for a different type of centrality (brokerage) than the closeness (proximity). Therefore, it gives very different results. The betweenness looks for vertices connecting separated subgroups. Granovetter (1973) states that such ties can be loose and yet very useful: Nodes with high betweenness have easy access to new information and they can benefit from the power of brokerage. In contrast, the ewc detects nodes having very close friends who often know each other as well. Therefore, the vertices with the highest and third highest ewc have a betweenness of zero, because they have no brokerage power.

The edge-weighted PageRank (Brin and Page 1998; Kiss and Bichler 2007) gives similar results as the ewc, but it emphasizes the line values implicitly. Therefore, the ranks of the third and fourth highest ewc nodes is switched by the weighted PageRank.

5 Analysis of a Telecommunication Network

Figure 6 shows an Ego’s 3-step calls graph from an Austrian telecommunications network. The line values represent call durations of two nodes within one month. The broader and darker the lines, the longer two people talked to each other. The size of the red vertices symbolize Freeman’s closeness and the size of the blue nodes indicate the ewc. In order to compare the two measures, the vertex size represents the rank, not the actual centrality score. So the largest blue vertex has the highest ewc and the second largest red node has the second highest closeness, for example.

Vertex A has many weak ties. This causes its closeness and its ewc to be approximately equally high, because the ewc allows a large number of loose friends to compensate close friendships. In contrast, Vertices B and C have relatively few friends, but they have a very strong relationship to each other. Consequently, their ewc is higher than their closeness. The group of blue nodes in the bottom right corner is in periphery of the network. Therefore their closeness is low. However, their line values and ewc3 shows that they are well connected locally. In contrast, the group of red vertices at the bottom left corner are closely connected to Vertex A, that is very central. Thus, their closeness is high while the small line values causes their ewc to be low.
6 Discussion

In order to create a weighted version of closeness that yields the intuitively correct results, we introduced the “edge-weighted closeness” (ewc), which incorporates line values as well as distances. We found out, that the ewc is somewhat the counterpiece of betweenness, because it identifies nodes having a circle of very good friends (which is not the idea of Freeman’s closeness). The ewc has similarities with the weighted degree and with the edge-weighted PageRank. However, the interpretation of the graphs clearly shows the differences. The telecommunications graph exemplary shows the differences between Freeman’s closeness and ewc and the derivable insights into the social behavior of nodes. The R code (Appendix A) code allows everyone to calculate the ewc.

The current definition and the R code allows the calculation of the ewc for connected graphs only. The question of how to calculate the closeness for disconnected graphs is as old as the closeness itself and not yet finally solved. As the line values increase the complexity of the problem, some solutions found for the closeness cannot be carried over to the ewc. This also means that the ewc only works for directed graphs if the network is still connected when only one
direction is considered.
In some applications, the line values do not correspond to a high proximity but
to great distance, like in traffic networks. A corresponding ewc version is in
progress.
Due to the large number of matrix multiplications necessary in the current code
(Appendix A), the calculation might take long for large networks. We tested
the ewc code for a graph with 3432 nodes and 9172 undirected lines on a dual
core 2.2 GHz machine with 2GB RAM. It took around 4.4 minutes on average.
As we have two variables determining the ewc, line values and distances, the
question of weighting these variables automatically arises. We are working on a
way to emphasize either the values, so that strong ties contribute more to close-
ness than many loose ties, or the distances, so that a large k-step neighborhood
becomes important.
Finally, we discovered that it can be interesting to see at which distance the
vertices gain most of their closeness. The most central node always gets most of
its closeness in the first distance. This does not have to be true for peripheral
vertices. Depending on the line values and the degree distribution, they might
get most of the closeness in the first distance or in higher distances. This insight
allows to group vertices according to their social roles and it is a topic a further
research.

A EWC Code for R

```r
### Edge-Weighted Closeness for R by Angela Bohn & Norbert Walchhofer ###

ewc <- function(vnetwork){
  network <- vnetwork
  network[network>0] <- 1
  W <- network
  Sigma <- network
  wd <- colSums(vnetwork, na.rm=TRUE)
  i <- 2
  while(min(Sigma)==0){
    Y <- vnetwork %*% W
    X <- network %*% W
    TM <- X
    TM[TM>0] <- 1
    TM <- TM-Sigma
    TM[TM<0] <- 0
    X <- TM * X
    diag(Y) <- 0
    Z <- (Y/X)/i
    Z[Z == "Inf" | Z == "-Inf"] <- 0
    wd <- rbind(wd,colSums(Z, na.rm=TRUE))
    Sigma <- Sigma + TM
    W <- W %*% network
    i <- i + 1
  }
  wd <- wd/(max(vnetwork)*(n-1))
  print(colSums(wd))
}```
Example:

```r
vnetwork <- cbind(c(0,1,0),c(1,0,10),c(0,10,0))
ewc(vnetwork)
[1] 0.300 0.550 0.525
```

The code can be downloaded from [http://www.angela-bohn.de/publications/ewc_code.html](http://www.angela-bohn.de/publications/ewc_code.html)

**References**


