

On the Betweenness Centrality of Trees, Forests and Cycles

Thomas Britz

School of Mathematics and Statistics, UNSW Sydney, NSW 2052, Australia, Email: britz@unsw.edu.au

ABSTRACT

Simple mathematical expressions are given for the betweenness centrality of nodes in trees, forests and cycles. As application, a centrality test is given for when a network might be a forest.

Keywords: Tree; Forest; Cycle; Network; Betweenness Centrality; Connected Component

1. Introduction

The advent of online social networks has made mathematical network theory intuitively accessible to non-mathematicians and this has lowered threshold barriers to the use of network theory in numerous areas of research. Now, network theory is used ubiquitously in chemistry, computer science, criminology, ecology, linguistics, logistics, medical big data, social sciences, telecommunications and many other areas; for a small representative sample, see^[1–8] for a tiny selection. Contemporary application of network theory often involves the measuring of centrality metrics. These can be used to help determine the topology of a network, aspects of which include particularly dense subcomponents and nodes or subcomponents with high center of gravity.

A particularly useful centrality metric is the betweenness centrality, first defined in^[9]. For a network G on nodes V , the betweenness centrality of any node v in G is defined by the sum

$$b_G(v) = \sum_{\{s,t\} \subseteq V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the number of shortest paths between node s and node t and $\sigma_{st}(v)$ is the number of such paths that pass through the node $v \neq s, t$.

The purpose of this note is to provide simple mathematical expressions for the betweenness centrality of nodes in trees, forests and cycles; see Theorem 1, 2 and 3 below. As an application of these results, a centrality test is given for when a network might be a forest; see Corollary 2.1.

2. Trees

To recall (see, e.g.,^[4]), a forest is, mathematically, a network with no cycle, and a tree is a connected forest. Trees can be graphically represented as collections of nodes and edges between the nodes, such as in the following illustration.

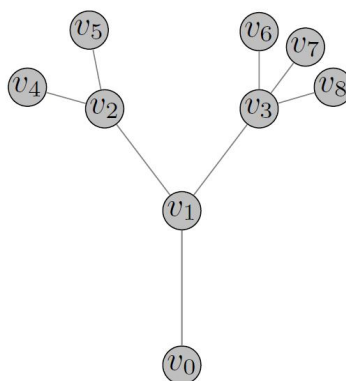


Figure 1; A tree T .

The general definition of betweenness centrality simplifies nicely in the case of trees T , namely to the simple expression

$$b_T(v) = \sum_{\{s,t\} \subseteq V} \sigma_{st}(v)$$

since there is exactly $\sigma_{st} = 1$ path between node s and node t . The betweenness centralities $b_T(v)$ of the nodes v for the tree T pictured in **Figure 1** are as given in the table below.

v	v_0	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
$b_T(v)$	0	19	13	18	0	0	0	0	0

Table 1. The betweenness centralities $b_T(v)$ of tree T .

Another expression for the betweenness centrality $b_T(v)$ of any node v of a tree T is presented and proved here, in Theorem 1 below. The importance of this theorem is that it allows $b_T(v)$ to be calculated effectively without the need for finding all shortest paths through v . Instead, it suffices to consider the sizes of the connected components of the network $T - v$ obtained by deleting from T the node v and each of its incident edges.

Theorem 1. For any node v of a tree T , let C_1, \dots, C_d denote the $d := \deg(v)$ connected components of $T - v$, and let c_1, \dots, c_d denote the respective number of nodes in these components. Then

$$b_T(v) = \sum_{1 \leq i < j \leq d} c_i c_j.$$

To illustrate the use of Theorem 1, consider the node v_1 in the tree T of Figure 1. This node has degree $d = \deg(v_1) = 3$, so the network $T - v_1$ has three connected components, say C_1 , C_2 , and C_3 , as depicted in Figure 2. These components contain $c_1 = 1$ nodes, $c_2 = 3$ nodes, and $c_3 = 4$ nodes, respectively. Therefore, by Theorem 1, the betweenness centrality $b_T(v_1)$ is

$$\begin{aligned} b_T(v_1) &= \sum_{1 \leq i < j \leq d} c_i c_j \\ &= c_1 c_2 + c_1 c_3 + c_2 c_3 \\ &= 1 \times 3 + 1 \times 4 + 3 \times 4 \\ &= 19. \end{aligned}$$

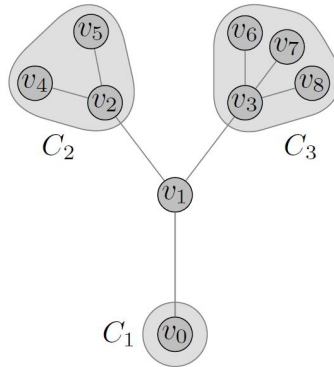


Figure 2. The connected components of $T - v_1$.

Proof of Theorem 1. Any path in T that passes through node v has endpoints s and t in distinct connected components of $T - v$, say C_i and C_j . The c_i nodes of C_i can be freely chosen as the endpoint in C_i ; similarly, the c_j nodes of C_j can be freely chosen as the endpoint in C_j . Thus, there are $c_i c_j$ choices for such endpoints. Summing over all pairs of connected components C_i and C_j therefore provides the betweenness centrality of v :

$$b_T(v) = \sum_{\{s,t\} \subseteq V} \sigma_{st}(v) = \sum_{1 \leq i < j \leq d} c_i c_j.$$

Note that if v is a leaf node, then $T - v$ has at most one connected component, in which case, then sum given by the theorem is empty and has value 0, which is indeed the value of $b_T(v)$.

A special case of Theorem 2 is that in which T is a path :

Corollary 1. For any node v of a path P , let i denote the closest distance from v to either end of P . Then

$$b_P(v) = i(n - 1 - i) .$$

3. Forests

We can extend Theorem 1 to forests as follows.

Theorem 2. For any node v of a forest G , let T denote the tree in G that contains v , let C_1, \dots, C_d denote the $d := \text{deg}(v)$ connected components of $T - v$, and let c_1, \dots, c_d denote the respective number of nodes in these components. Then

$$b_G(v) = b_T(v) = \sum_{1 \leq i < j \leq d} c_i c_j .$$

From Theorem 2, it is clear that the betweenness centrality of each node of a tree, or a forest more generally, must be an integer. This provides us with a simple centrality test for testing whether a network is a forest, as follows.

Corollary 2.1. If the betweenness centrality $b_G(v)$ of any node v of a network G is not an integer, then G is not a forest.

To demonstrate this test, consider the network G in **Figure 3** below. This network is a cycle and is therefore not a forest. This is also evident from the betweenness centralities

$$b_G(v_1) = b_G(v_2) = b_G(v_3) = b_G(v_4) = \frac{1}{2} .$$

Since these betweenness centralities are not all integers, Corollary 2 asserts that G is not a forest.

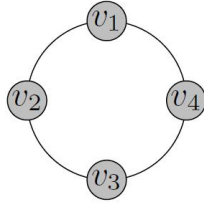


Figure 3; A non-forest network G .

4. Cycles

As the example above demonstrated, the nodes of a cycle network can have non-integer betweenness centralities. However, cycles are very nearly trees, and indeed, it turns out that in the majority (75%) of cases, the node betweenness centralities of a cycle are in fact integers; see Remark 3.1 below. This is implied by the following Theorem 3, in which the node betweenness centralities of the cycle C_n with n nodes is given explicitly in terms of the value of n .

Theorem 3. For any node v of the cycle C_n with n nodes.

If n is even, then

$$b_{C_n}(v) = \frac{1}{8}(n - 1)(n - 3) .$$

If n is odd, then

$$b_{C_n}(v) = \frac{1}{8}(n - 2)^2 .$$

To illustrate Theorem 3, consider the network G in **Figure 3**. It is the cycle C_4 so, since 4 is even, each node in G has betweenness centrality

$$\frac{1}{8}(n - 2)^2 = \frac{1}{8}(4 - 2)^2 = \frac{1}{2} ,$$

as previously established.

Note that if n is odd, then both $n-1$ and $n-3$ are even, and precisely one of these is divisible by 4; therefore, $b_{C_n}(v) = \frac{1}{8}(n-1)(n-3)$ is an integer. If n is even, however, then $b_{C_n}(v) = \frac{1}{8}(n-2)^2$ may or may not be an integer. In particular, this expression is even and thus an integer if $n-2$ is divisible by 4; however, it is not an integer if n is divisible by 4. We may summarise these remarks as follows.

Remark 3.1. For any node v of the cycle C_n with n nodes, $b_{C_n}(v)$ is an integer unless n is divisible by 4.

Proof of Theorem 3. Label the nodes of C_n consecutively by $0, 1, \dots, n-1$, visualised as lying clockwise around a circle, and note that, by symmetry, the node betweenness centralities are all equal. It therefore suffices to calculate $b_{C_n}(v)$ for the node $v = 0$.

We must consider two cases, namely that in which n is even and that in which n is odd. The detail that distinguishes these cases is that, in the even case, σ_{st} can equal 2, namely when nodes s and t lie on diametrically opposite sides of C_n ; in other words, when the difference between the label values of s and t is $\frac{n}{2}$, for then there are exactly 2 shortest paths between s and t . In the odd case, no pair of nodes lie diametrically opposite each other, so $\sigma_{st} = 1$ for all nodes s and t . This is illustrated in **Figure 4**.

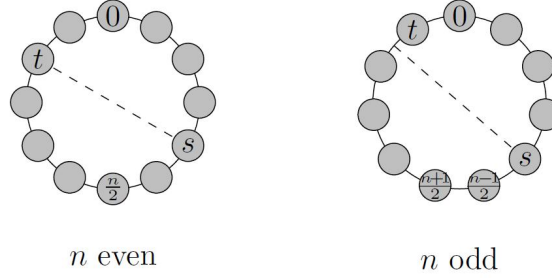


Figure 4; The cycle C_n .

Consider the case in which n is even, and suppose that s is one of the nodes $1, \dots, \frac{n}{2} - 1$. If $t = s + \frac{n}{2}$, then s and t lie on diametrically opposite sides of C_n , and $\sigma_{st} = 2$ and $\sigma_{st}(v) = 1$. If $t \in \{s + \frac{n}{2} + 1, \dots, n-1\}$, then there is a unique shortest path between s and t , and it passes via the node $v = 0$, so $\sigma_{st} = \sigma_{st}(v) = 1$. For the values $t = s + 1, \dots, s + \frac{n}{2} - 1$, the shortest path between s and t does not pass through v , so $\sigma_{st}(v) = 0$.

Therefore,

$$\begin{aligned}
 b_{C_n}(v) &= \sum_{\{s,t\} \subseteq \{0, \dots, n-1\}} \frac{\sigma_{st}(v)}{\sigma_{st}} \\
 &= \sum_{s=1}^{\frac{n}{2}-1} \left(\frac{1}{2} + \sum_{t=s+\frac{n}{2}+1}^{n-1} 1 \right) \\
 &= \sum_{s=1}^{\frac{n}{2}-1} \left(\frac{1}{2} + n - 1 - \left(s + \frac{n}{2} \right) \right) \\
 &= \binom{n}{2} \left(\frac{n}{2} - \frac{1}{2} \right) - \sum_{s=1}^{\frac{n}{2}-1} s \\
 &= \left(\frac{n}{2} - 1 \right) \left(\frac{n}{2} - \frac{1}{2} \right) - \frac{1}{2} \left(\frac{n}{2} - 1 \right) \frac{n}{2} \\
 &= \frac{1}{8} (n-2)^2.
 \end{aligned}$$

The odd case is similar.

References

1. O. Abramov, Network Theory Applied to Linguistics - New Advances in Language Classification and Typology, PhD thesis, Bielefeld University, 2011.
2. R. Aris, "Prolegomena to the rational analysis of systems of chemical reactions," Archive for Rational Mechanics and Analysis, vol. 19, pp. 81-99, 1965.
3. D. Bright, C. Greenhill, T. Britz, A. Ritter and C. Morselli, "Criminal network vulnerabilities and adaptations," Global Crime, vol. 18, pp. 424-441, 2017.
4. R. Diestel, Graph Theory, 2nd ed., Springer-Verlag, 1999.
5. L. Ford, Jr. and D. Fulkerson, Flows in Networks, Princeton University Press, 1962.
6. D. Hunter, T. Britz, M. Jones and M. Letnic, "Reintroduction of Tasmanian devils to mainland Australia can restore top-down control in ecosystems where dingoes have been extirpated," Conservation Biology, vol. 191, pp. 428-435, 2015.

7. E. Otte and R. Rousseau, "Social network analysis: a powerful strategy, also for the information sciences," *Journal of Information Science*, vol. 28, pp. 441-453, 2002.
8. B. Tran, P. Straka, M. Falster, K. Douglas, T. Britz and L. Gorm, "Overcoming the data drought: exploring general practice in Australia by network analysis of big data," *Medical Journal of Australia*, Vols. , to appear, 2018.
9. L. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, pp. 35-41, 1977.