MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **68** (2012) 293-302

ISSN 0340 - 6253

Centrality and Betweenness: Vertex and Edge Decomposition of the Wiener Index

Gilles Caporossi[a,*], Marcia Paiva[b], Damir Vukičević[c] and Marcelo Segatto[d]

[a] *GERAD, HEC Montréal, Montréal, Canada*

[b] *GERAD and Federal University of Espírito Santo, Vitória, Brazil*

[c] *Department of Mathematics, University of Split, Split, Croatia*

[d] *Federal University of Espírito Santo, Vitória, Brazil*

(Received October 3, 2011)

Abstract

In this paper, we present an edge and vertex decomposition of the Wiener index (W) that is related to the concept of betweenness centrality used in social networks studies. Some classical methods to compute W could easily be derived from this formulation and novel invariants may be defined by this mean. Another vertex decomposition of W is the transmission. If transmission and centrality are both vertex decompositions of W it seems that they are represent opposite concepts, however the nature of this relation is not always so clear. Some properties obtained with the AutoGraphiX software on betweenness, centrality and their relation to transmission are presented and proved.

1 Introduction

In 1947 [1], H. Wiener proposed the now so called Wiener index, W. The Wiener index is certainly one of the most studied in mathematical chemistry and graph theory. In its original version, W was only defined for trees $T = (V, E)$ as follows:

$$\sum_{(i,j)\in E} n_i \cdot n_j \tag{1}$$

where n_i is the number of vertices that are closer to i than j and vice versa.

*Gilles.Caporossi@gerad.ca

If we note $G = (V, E)$ a simple connected graph of order $n = |V|$ and size $m = |E|$, in which d_{ij} is the distance between vertices i and j (number of edges in a shortest path between i and j),

$$W = \sum_{i=1}^{n} \sum_{j=1}^{i} d_{ij}. \tag{2}$$

As stressed by Dobrynin and Gutman [2], when applying equation 1 to a general graph, one gets a value W^* that may get larger than W. Conditions to have $W^* = W$ are exposed in [2]. The index W^* was later called Szeged index (Sz) and has also been widely studied. If shortest paths between pairs of vertices are unique, $Sz = W$ and computing W could be achieved by decomposition of W by edges. If it is not the case, choosing one of the shortest paths at random will provide a way to compute W. This property is implicitly used in the cuts decompositions of W. In the present paper, we are not interested in finding a way to compute W but we propose to study a decomposition of W as a way to define values associated to vertices or edges. For such a study, the contribution of each edge or vertex must be unique and well defined. In a completely different field, Freeman [3] proposed the concept of betweenness as an evaluation of the centrality for studying social networks. In the original definition, betweenness centrality of a vertex v is defined by the probabilistic number of shortest paths between pairs of vertices that uses vertex v. In the present paper, we use a slightly different way to look at betweenness in the sense that it is first defined for edges and later extended to vertices by summation. Each edge $(i, j) \in E$ is thus associated the probabilistic number of shortest paths b_{ij} in which it is involved as follows:

$$b_{ij} = \sum_{k,l \in V, k<l} \frac{s_{ij}^{kl}}{s^{kl}} \; \forall (i, j) \in E. \tag{3}$$

Where s_{ij}^{kl} is the number of shortest paths between vertices k and l ($k < l$) that uses the edge (i, j) and s^{kl} is the total number of shortest paths between k and l. From this definition, a centrality measure for vertices may be defined as follows:

$$C_i^* = \sum_{j \in V, (i,j) \in E} b_{ij} \; \forall i \in V. \tag{4}$$

This measure is very close, yet different to Freeman's betweenness centrality C_i. Namely, we have $C_i^* = 2C_i + n - 1$. Given the relation between these two quantities, C_i^* is called adjusted betweenness centrality.

It turns out that the summation of b_{ij} over all edges of the graph equals $2W$, as well as the summation of C_i^* over all vertices, which makes b_{ij} and C_i^* ways to decompose

W by edges and vertices. Another well known way to decompose W by vertices is the transmission T_i which is defined as follows [4]:

$$T_i = \sum_{j \in V} d_{ij}. \tag{5}$$

In this paper, we propose to study the edge and vertex decomposition of the Wiener index, and identify some properties of the so found edge or vertex values.

2 Edge decomposition of W

There are various ways to assign weights to edges of a graph so that they sum up to W. In order to make sense, it must respect some rules. The first is that for a given pair of vertices ij, the sum of the weights of edges in shortest paths between i and j sums up to the distance between i and j. A way is thus to choose randomly one shortest path and assign it a weight of 1. This way involves a random choice that brings instability to the values obtained. Another way consists in assigning an equal weight to each shortest path between i and j as proposed by [5] and [6]. This last way to weight edges of G is much better because it could be considered as an average weight if the probability to use any shortest path is equal. This interpretation could stand for random connections between pairs of vertices in the context of a telecommunication network, it could be considered as a centrality indicator in the case of social networks and is certainly more accurate than the other in the case of chemical graphs.

2.1 Weight of the edges of a cocycle

In this section, we consider partitions of the graph $G = (V, E)$ into $G_A = (V_A, E_A)$ and $G_B = (V_B, E_B)$ such that G_A and G_B are connected. We denote C_{AB} the cocycle $E \backslash (E_A \cup E_B)$.

Property 1 *Let G_A, G_B be a partition of a graph G and denote by C_{AB} the set of edges between vertices of G_A and G_B.*

$$\sum_{e \in C_{AB}} b_e \geq |V_A| \cdot |V_B| \tag{6}$$

and the inequality is strict if and only if there is at least a pair of vertices of G_A (or G_B) having a shortest path going thru a vertex of G_B (or G_A respectively). If the equality holds, we say that we found a valid cocycle.

Proof.

Any shortest path from a vertex $a \in V_A$ and a vertex $b \in V_B$ contains at least an edge in C_{AB}. It follows immediately that the number of shortest paths crossing C_{AB} is at least $|V_A| \cdot |V_B|$.

Lemma 1 *Let V_A, V_B be a partition of the vertices V of G and C_{AB} the corresponding cocycle. We have*

$$\sum_{e \in C_{AB}} b_e = |V_A| \cdot |V_B| \tag{7}$$

if and only if for any pair of vertices a_1 and a_2 of V_A and b_1, b_2 of V_B such that a_1 is adjacent to b_1 and a_2 to b_2, eventually, $b_1 = b_2$, the distances d_a between a_1 and a_2 and d_b between b_1 and b_2 will not differ by more than 1. In this case, we say that C_{AB} is a valid cut.

Proof.

We first note that the equality from equation 7 will hold if no shortest path containing an edge of C_{AB} will join two vertices from the same subgraph because b_{ij} is the summation of non negative terms and cannot be lower than $|V_A| \cdot |V_B|$ because this value is that of the number of shortest paths from vertices of G_A to vertices of G_B only. Suppose now by contradiction that $d_a \geq d_b + 2$, then the path composed of the edge (a_1, b_1) followed by the shortest path p_b between b_1 and b_2 and (b_2, a_2) will have the length $1 + d_b + 1 = d_b + 2 \leq d_a$, in which case the cut C_{AB} contains some shortest paths between vertices of V_A.

Corollary 1 *For any edge $(i, j) \in C_{AB}$, no vertex of V_A is closer to j than i and vice versa.*

Corollary 2 *If $|C_{AB}|$ is a valid cut, the subgraph G_A (G_B) induced by V_A (V_B) is connected.*

We note that a partition of the edges of G in valid cocycles will provide the now classical method to compute W for benzenoids [7, 8, 9, 10] and other classes of graphs [11, 12].

2.2 Graphs with extremal values for b_{ij}

Theorem 1 *Let* $b^{max} = \max_{(ij)\in E} b_{ij}$ *be the largest betweenness of* G*, then we have*

$$1 \le b^{max} \le \left\lfloor \frac{n^2}{4} \right\rfloor . \tag{8}$$

The lower bound is tight if and only if G *is a complete graph and the upper bound is tight if and only is* G *is composed of two subgraphs* G_A *and* G_B *of respective order* a *and* b *such that* $a = \lfloor \frac{n}{2} \rfloor$ *and* $b = \lceil \frac{n}{2} \rceil$*, joined by a single edge.*

Proof.

The proof for the lower bound is obvious as the complete graph is the only graph for which all edges have betweenness 1, which is the minimal value possible for betweenness. Any graph which has at least two non adjacent vertices i and j has higher b^{max} because the shortest path between i and j will increase betweenness of some edges. The proof for the upper bound is the following: For a given edge (i, j), let V_i be the set of vertices closer to i than j and V_j the set of edges closer to j. If a vertex is at equal distance from i and j, it may randomly be assigned to V_i or V_j. Let C_{ij} be the set of edges between vertices of V_i and V_j.

By corollary 2, the subgraphs induced by V_i and V_j are both connected, thus it is possible to remove all edges of C_{ij} except (i, j). In this case b_{ij} is increased to $|V_i| \times |V_j|$. If we set $a = |V_i|$, we have $b_{ij} = a \times (n - a)$ that is maximized for $a = \lfloor \frac{n}{2} \rfloor$. □

An equivalent formula for the upper bound could be found in [13] but no formal proof was provided.

Theorem 2 *Let* $b^{\min} = \min_{(i,j)\in E} b_{ij}$ *be the smallest betweenness of* G*, then we have*

$$1 \le b^{\min} \le \lfloor n^2/4 \rfloor .$$

The lower bound is reached for any edge of the complete graph and upper bound is reached for any edge of the cycle.

Proof.

The lower bound is obvious. Let us prove the upper bound. Suppose to the contrary. Let G be a graph such that $b^{\min} = b^{\min}(G) > \lfloor n^2/4 \rfloor$. If G has a pendant vertex, then $b^{\min} = n - 1$ and the claim holds since $n - 1 \le n^2/4$ and $n - 1$ is an integer. Hence, let us suppose that G has no pendant vertices. Let us distinguish two cases:

1. G is biconnected. Then, transmission of each vertex is at most $\lfloor n^2/4 \rfloor$ as stated by Plesnik [14] and the number of edges is larger than number of vertices, hence $b^{\min} \leq \lfloor n^2/4 \rfloor$.

2. G is not biconnected.

   Note that there are at least two biconnected components (blocks) that are incident to one cut vertex. Let us denote them by V_1 and V_2 and corresponding cut vertices by c_1 and c_2. Let v_i be the vertex in V_i on the greatest distance from c_i and (v_i, w_i) an edge in V_i incident to $v_i, i = 1, 2$. Further, let us denote by x_1 and x_2 cardinalities of these blocks where cut vertices are not taken into the consideration. It holds that

$$b_{v_i w_i} \leq (n - x_i - 1) + \binom{x_i}{2},$$

   where the first bracket corresponds to the shortest path connecting vertex v_i with some vertex outside $V_i \backslash \{c_i\}$ and the binomial coefficient corresponds to the shortest paths connecting vertices within $V_i \backslash \{c_i\}$. It holds

$$(n - x_i - 1) + \binom{x_i}{2} \geq \frac{n^2 - 1}{4}.$$

   It implies that

$$x_i \geq \frac{1}{2}\left(3 + \sqrt{15 - 8n + 2n^2}\right) \geq \frac{1}{2}\left(3 + \sqrt{2}\,(n-2)\right) \geq \frac{\sqrt{2}}{2} n,$$

   but then $x_1 + x_2 > n$ which is a contradiction. □

3 Centrality and Transmission: vertex decompositions of W

As stated before, the adjusted betweenness centrality C_i^* of a vertex i is defined as follows:

$$C_i^* = \sum_{j/(i,j) \in E} b_{ij}, \tag{9}$$

and the transmission T_i as follows:

$$T_i = \sum_{j \in V} d_{ij}, \tag{10}$$

where d_{ij} is the distance between vertex i and j. If we have

$$\sum_{i \in V} C_i^* = \sum_{i \in V} T_i = 2W, \tag{11}$$

both quantities have different behaviors.

3.1 Relation between T_i and C_i

Intuitively, central vertices tend to have lower transmission and larger centrality while it is the opposite when a vertex gets to the border. A natural question is to know wether centrality and transmission are correlated and wether they both indicates the same central vertex, i.e. is it always true that the vertex minimizing transmission is the one that maximizes centrality. Unfortunately, this relation does not hold, even for trees as shows the following counter-example where vertex E minimizes transmission (17) while the centrality is maximized by F (44).

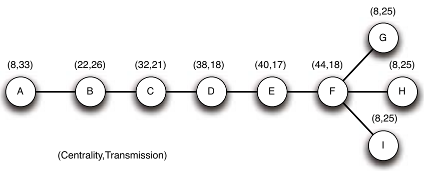

Figure 1: A tree for which the vertex maximizing centrality is not minimizing transmission.

3.2 New results on centrality

Based upon some known results, we determine some basic relations obtained with AutoGraphiX [15].

Theorem 3 *Let* $C^{max} = \max_{i \in V} C_i$ *denote the maximum centrality of a graph* G *of order* n*, then we have*

$$n - 1 \leq C^{max} \leq (n-1)^2. \tag{12}$$

The lower bound is reached by any vertex of a complete graph and the upper bound for the central vertex of a star.

The lower bound can easily be proved as the minimum value of C_i is $n-1$ and this is also the maximum value for complete graphs. The upper bound can also easily be proved because the star is the only configuration for which all shortest paths between pairs of vertices involve a common vertex.

Theorem 4 *Let $C^{min} = \min_{i \in V} C_i$ denote the minimum centrality of a graph G of order n, then we have*

$$n-1 \leq C^{min} \leq \begin{cases} \dfrac{n^2}{4} & n \text{ even;} \\ \dfrac{n^2-1}{4} & n \text{ odd.} \end{cases}$$

The lower bound is reached by vertices whose neighbors are all adjacent and the upper bound is reached by any vertex in a cycle.

Proof.

The proof of the lower bound is rather simple as each vertex is involved in at least $n-1$ shortest paths, those from it to other vertices. To prove the upper bound, we will consider two cases:

1. G is biconnected with respect to the vertices.

   In this case, the transmission of each vertex is at most $\lfloor \frac{n^2}{4} \rfloor$ as stated by Plesnik [14] moreover this bound is reached if and only if G is a cycle. From this relation, we deduce that

$$\bar{T}_i = \bar{C}_i = \begin{cases} \dfrac{n^2}{4} & n \text{ even;} \\ \dfrac{n^2-1}{4} & n \text{ odd.} \end{cases}$$

   In the case of a cycle, $C_i = \lfloor \frac{n^2}{4} \rfloor \ \forall i$. Thus the theorem is true for biconnected graphs and the bound is tight for cycles.

2. G is not biconnected.

   Let G be a graph with $n \geq 5$ vertices and at least one cut vertex. Then,

$$C^{min} \leq \begin{cases} \frac{n^2}{4} & n \text{ even;} \\ \frac{n^2-1}{4} & n \text{ odd.} \end{cases}$$

   Suppose to the contrary. Let G be a such graph. If G has a pendant vertex, then $C^{min} = n-1$ and the claim holds. Hence, let us suppose that G has no pendant vertices. Note that there are at least two bi-connected components (blocks) that are incident to one cut vertex. Let us denote them by C_1 and C_2 and corresponding cut vertices by c_1 and c_2. Let v_i be the vertex in C_i on the greatest distance from c_i, $i = 1, 2$. Further, let us denote by x_1 and x_2 cardinalities of these blocks where cut vertex is not taken into the consideration. It holds that

$$c(v_1) \leq (n-1) + \left(2 \cdot \binom{x_1 - 1}{2}\right),$$

where the expression in the first bracket is the sum of contributions to $c(v_1)$ of all shortest paths starting in v_1. Further, note that any other shortest path contributing to $c(v_1)$ has to have both end-vertices in $C_1 \setminus \{c_1\}$, because v_1 is the vertex furthest from c_1. Since each set of paths between such two vertices can contribute 2 to edges incident to v_1, it follows that its total contribution is at most $2 \cdot \binom{x_1 - 1}{2}$. Hence

$$(n-1) + \left(2 \cdot \binom{x_1 - 1}{2}\right) \leq \frac{n^2 - 1}{4}.$$

It implies $x_1 \geq (n+1)/2$. Completly analogously, it can be shown that $x_2 \geq (n+1)/2$. But, then $x_1 + x_2 > n$, which is a contradiction.

Theorem 5 *Let $T^{\min} = \min_{i \in V} T_i$ denote the minimum transmition, then we have $C^{\min} \geq T^{\min} - \left\lfloor \frac{n^2}{4} \right\rfloor + n - 1$ and the equality is tight for paths.*

Proof.
It can be easily checked that $T^{\min}(P_n) - C^{\min}(P_n) = \left\lfloor \frac{n^2}{4} \right\rfloor - n + 1$. Hence, it is sufficent to prove that $T^{\min}(G) - C^{\min}(G) \leq T^{\min}(P_n) - C^{\min}(P_n)$ for every connected graph G with n vertices. Moreover, from Theorem 3, it follows that it is sufficient to prove that $T^{\min}(G) \leq T^{\min}(P_n)$ and this is proved in [16].

Acknowledgments

The authors would like to thank Snježana Majstorović for her useful comments. This work was supported by CAPES and FAPES (Brazil), NSERC (Canada) and Croatian Ministry of Sciences (Grants nos. 177/0000000/0884 and 037/0000000/2779) and project GReGAS within EuroGIGA collaborative research project.

References

[1] H. Wiener, Structural determination of paraffin boiling points, *J. Amer. Chem. Soc.* **69** (1947) 17–20.

[2] A. Dobrynin, I. Gutman, On a graph invariant related to the sum of all distances in a graph, *Publ. Inst. Math.* **56** (1995) 18–22.

[3] L. C. Freeman, A set of measures of centrality based on betweenness, *Sociometry* **40** (1977) 35–41.

[4] B. Mohar, T. Pisanski, How to compute the Wiener index of a graph, *J. Math. Chem.* **2** (1988) 267–277.

[5] I. Lukovits, Correlation between components of the Wiener index and partition coefficients of hydrocarbons, *Int. J. Quantum Chem.* **19** (1992) 217–223.

[6] M. Juvan, B. Mohar, Bond contribution to the Wiener index, *J. Chem. Inf. Comput. Sci.* **35** (1995) 217–219.

[7] I. Gutman, S. Klavžar, A method for calculating Wiener numbers of benzenoid hydrocarbons and phenylenes, *ACH Models Chem.* **133** (1996) 389–399.

[8] I. Gutman, Comparing Wiener numbers of isomeric benzenoid hydrocarbons, *ACH Models Chem.* **134** (1997) 477–486.

[9] I. Gutman, Wiener numbers of benzenoid isomers containing a linear polyacene fragment, *Indian J. Chem.* **36A** (1997) 644–648.

[10] A. Dobrynin, I. Gutman, S. Klavžar, P. Žigert, Wiener Index of Hexagonal Systems, *Acta Appl. Math.* **72** (2002) 247–294.

[11] W. C. Shiu, P. C. B. Lam, K. K. Poon, On Wiener numbers of polygonal nets, *Discr. Appl. Math.* **122** (2002) 251–261.

[12] S. Klavžar, A bird's eye view of the cut method and a survey of its applications in chemical graph theory, *MATCH Commun. Math. Comput. Chem.* **60** (2008) 255–274.

[13] F. Comellas, S. Gago, Spectral bounds for the betweenness of a graph, *Lin. Algebra Appl.* **423** (2006) 74–80.

[14] J. Plesnik, On the sum of all distances in a graph or digraph, *J. Graph Theory* **8** (1984) 1–21.

[15] G. Caporossi, P. Hansen, Variable neighborhood search for extremal graphs. I: The AutoGraphiX system, *Discr. Math.* **212** (2000) 29–44.

[16] M. Aouchiche, P. Hansen, Proximity and remoteness in graphs: Results and conjectures, *Networks*, in press.