7-2020

Unpacking Burt’s Constraint Measure

Martin G. Everett
University of Manchester, UK

Stephen P. Borgatti
University of Kentucky, sborgatti@uky.edu

Digital Object Identifier (DOI)
https://doi.org/10.1016/j.socnet.2020.02.001

Notes/Citation Information
Published in *Social Networks*, v. 62.

Unpacking Burt's Constraint Measure

Martin G Everett*
The University of Manchester

Stephen P Borgatti
University of Kentucky

*Mitchell Centre for Social Network Analysis
School of Social Sciences
University of Manchester
Arthur Lewis Building
Bridgeford Street
Manchester
M13 9PL
UK
e-mail martin.everett@manchester.ac.uk
Tel +44 (0) 161 275 2515

Abstract

Burt (1992) proposed two principal measures of structural holes, effective size and constraint. However, the formulas describing the measures are somewhat opaque and have led to a certain amount of confusion. Borgatti (1997) showed that, for binary data, the effective size formula could be written very simply as degree (ego network size) minus average degree of alters within the ego network. The present paper presents an analogous reformulation of the constraint measure. We also derive minima and maxima for constraint, showing that, for small ego networks, constraint can be larger than one, and for larger ego networks, constraint cannot get as large as one. We also show that for networks with more than seven alters, the maximum constraint does not occur in a maximally dense or closed network, but rather in a relatively sparse "shadow ego network", which is a network that contains an alter (the shadow ego) that is connected to every other alter, and where no other alter-alter ties exist.

Key words: Egonetworks; structural holes; constraint.

Highlights

- An alternative formulation for constraint is derived when the data is undirected and binary
- A means of approximating constraint for undirected binary data when the alter-alter ties are unknown is presented
- We give minimum values for constraint and conjecture maximum values for both binary and valued data
- We show that, for larger ego networks, maximum constraint occurs not when an ego network is maximally dense, but when it is shaped as a shadow ego network

1. Introduction

Burt's (1992) theory and accompanying measures of structural holes have become a mainstay of the ego network literature, attracting tens of thousands of citations. Of Burt's measures, the most successful, both in terms of adoption by the field and proven ability to predict ego outcomes, is constraint. As a measure of structural holes, the constraint measure is intended to capture the extent to which an individual has access to many non-redundant others. The lack of a tie among a node's contacts represents opportunities for brokerage – for combining the knowledge and efforts of different individuals in such a way as to provide value to them, to the broker, and to others at large.

Burt's other well-known measure of structural holes is effective size. The formulas for both measures look daunting, but, as Borgatti (1997) has shown, for non-valued network data, the expression for effective size can be simplified considerably. Effective size is simply a person's ego network size (i.e., the number of contacts they have) minus the average number of ties each contact has within the ego network (not including ego). The latter term is known as average degree and is equal to the density of the ego network (excluding ties to ego) multiplied by N-1. Thus, network size contributes positively toward their structural holes score, while density/average degree contributes negatively toward structural holes. To maximize social capital, then, an actor would want to have many contacts who are unconnected to each other, other than indirectly through the broker.

The situation with constraint is different. The expression for constraint is more difficult to understand, and the measure is often misinterpreted or misused. For example, the measure is technically undefined for an isolate – a node with no alters. But as an anonymous reviewer of this paper pointed out, researchers not infrequently assign a zero for the node's constraint score. Constraint is a reverse measure of structural holes: the bigger the numerical value, the fewer the structural holes and the lower the social capital. A constraint score of zero would indicate an individual bursting with structural holes, who, according to the theory, would be expected to be a high performer. But, in reality the individual is an isolate with no access to social resources. Similarly, there is confusion about the range of the constraint scores. Most people assume that it is bounded from above by one, and think there must be a bug in the software when a value greater than one is obtained. Yet, as this paper will show, constraint is not bounded by one. Nor is the minimum zero. Another curiosity of the empirical literature using structural holes is the practice of controlling for network size when regressing an outcome on constraint. The problem is that size is a fundamental part of constraint and is an explicit part of its formulation. An ego network with fewer contacts has fewer structural holes and lower social capital, all else being equal. Controlling for size eviscerates constraint, leaving the variable labeled constraint in the regression a measure of something other than structural holes as Burt conceived them.

This paper is an attempt to clarify what the constraint measure is and does. The approach we take is similar to Borgatti (1997), in that we deliberately take the special case of non-valued undirected networks to simplify the measure and examine its underpinnings.

2. Constraint

Burt (1992) defines both overall constraint (ego-level) and dyadic constraint. Dyadic constraint, c_{ij}, is the extent to which actor j constrains actor i (whom we shall refer to as ego).

Overall constraint, c_i, is the sum over all j in i's neighbourhood.

$$c_i = \sum_{j \in N(i)} c_{ij}$$

Burt defines dyadic constraint c_{ij} as

$$c_{ij} = \left(p_{ij} + \sum_{q \in N(i)-j} p_{iq} p_{qj} \right)^2 \tag{1}$$

where p_{ij} is the amount of energy actor *i* invests in actor *j*. If we have a valued adjacency matrix A in which higher values represent a greater investment in energy then we construct P from A as follows.

$$p_{ij} = \frac{a_{ij} + a_{ji}}{\sum_j (a_{ij} + a_{ji})}$$

Since the total constraint is the sum of the constraint for each alter in ego's network. It follows that the constraint on ego *i* is

$$\sum_j c_{ij} = \sum_j \left(p_{ij} + \sum_q p_{iq} p_{qj} \right)^2$$

$$= \sum_j p_{ij}^2 + \sum_j 2p_{ij} \sum_q p_{iq} p_{qj} + \sum_j \left(\sum_q p_{iq} p_{qj} \right)^2 \tag{2}$$

This last formulation was discussed by Burt (1998, pg. 42, footnote 14). He describes the three components as follows: "The first variable in the expression, C-size in the text, is a Herfindal index measuring the extent to which manager i's relations are concentrated in a single contact. The second variable, C-density in the text, is an interaction between strong ties and density in the sense that it increases with the extent to which manager i's strongest relations are with contacts strongly tied to the other contacts. The third variable, C-hierarchy in the text, measures the extent to which manager i's contacts concentrate their relations in one central contact."

It is clear the first term only looks at ego-alter ties and therefore is largely influenced by network size. This term is large to the extent node *i* has a strong tie with j, and that i has few other strong ties). Both the second and third terms contain alter-alter ties and the number of such ties will clearly impact the exact relationship between the second term and density and the third term and hierarchy. The exact nature of these relations are not easy to deduce. In trying to unpack the constraint measure there are two major difficulties. The first is the fact that the edges are all valued and it is difficult to simplify or manipulate the constraint formula with valued data. Secondly, the formula contains a squared summation and this non-linearity gives an extra level of complexity.

3 Undirected Binary Data

In order to make progress in unravelling constraint, we simplify the problem by considering constraint on simple binary undirected graphs. In this case we know that ego will be connected to every alter by an undirected edge with value 1. To simplify the expressions, we remove the subscript *i*, since it is understood that it always refers to ego. We then use N to refer to the number of alters. To simplify some expressions at a later stage, we shall also assume that ego has been deleted from the ego network and we will just consider the network of alter-alter ties. In our ego-deleted network we shall denote the degree of alter *j* by $\rho(j)$. It now follows that the matrix P (the proportion of relational energy each actor invests in each contact) can be written in terms of each node's degree as $p_{ij}=1/N$ and $p_{qj}=1/(\rho(q)+1)$. We now substitute these values into Equation 2 to obtain

$$\sum_j \left(\frac{1}{N^2} + \frac{2}{N^2} \sum_q p_{qj} + \frac{1}{N^2} \left(\sum_q p_{qj} \right)^2 \right)$$

The first sum over j is over all the alters of i and therefore runs from 1 to N. Therefore, this expression can be written as

$$\frac{1}{N} + \frac{2}{N^2} \sum_j \sum_q p_{qj} + \frac{1}{N^2} \sum_j \left(\sum_q p_{qj} \right)^2$$

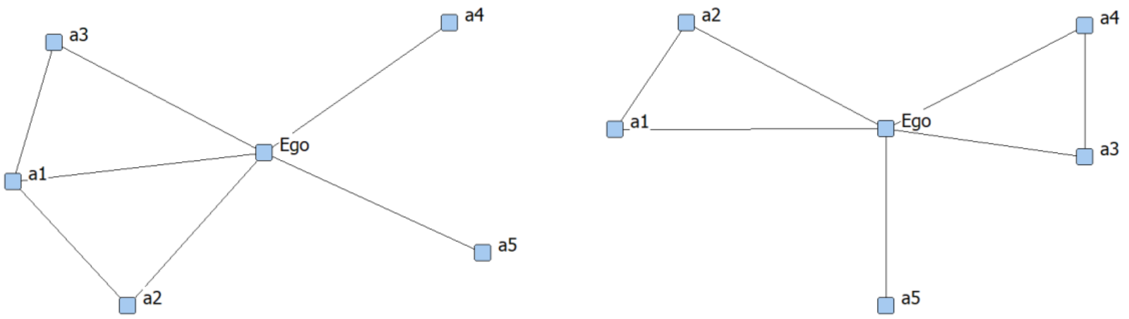
The second term in this new expression is the sum over *j* of the sum over *q* of p_{qj}, the sum over *q* is for actors in the neighbourhood of j. Consider all neighbours of *j* (again, not including ego). Then *j*'s contribution to the neighbours' constraint is the reciprocal of the degree of j, which is one more than the degree in the ego-deleted neighbourhood, as it must include ego, and hence is $1/(\rho(j)+1)$. But there are $\rho(j)$ of these alters so we can write the expression as

$$\frac{1}{N} + \frac{2}{N^2} \sum_j \frac{\rho(j)}{\rho(j)+1} + \frac{1}{N^2} \sum_j \left(\sum_q p_{qj} \right)^2$$

$$= \frac{1}{N} + \frac{2}{N^2} \sum_{j=1}^{N} \frac{\rho(j)}{\rho(j)+1} + \frac{1}{N^2} \sum_{j=1}^{N} \left(\sum_{q \epsilon N(j)} \frac{1}{\rho(q)+1} \right)^2 \quad (3)$$

Another way to understand what we have done with the second term is to consider how we calculate the constraint of each of the alters. In Burt's formulation we take each alter in turn and calculate how it is constrained by the actors it is connected to. In our formulation we take each alter in turn and calculate how much it constrains the actors it is connected to. Since these values are all summed in the outer sum, the result will be the same. In the first case we are summing a collection of different values for each of the alters and hence we have a double sum, in our re-arrangement the same single value can now be used as we are seeing how the chosen alter constrains those it is connected to. Since we now have a single value we can eliminate the inner sum as it is $1/(\rho(j)+1)$ counted $\rho(j)$ times. It should be noted that this is only possible because in the binary case each edge has the same value.

We can now see more clearly how the three terms contribute to the measure, at least for the binary case. The first term is just the reciprocal of network size. The larger the network, the smaller this term. The second term is dependent entirely on the degree sequence in the ego network. To calculate it, we do not need to know which alter has which degree -- a simple list of the degrees (the degree sequence) is sufficient. Two networks with the same degree sequence will have the same Term 2 score. Of course, having the same degree sequence implies having the same density. In this sense, we might say this term captures density. However, there is more to it. There are different degree sequences that can yield the same density, and these will have different scores on term 2. Figure 1 shows two ego networks with identical densities but different degree sequences. Specifically, the one on the left with the wider range of alter degrees has a higher score on term 2.

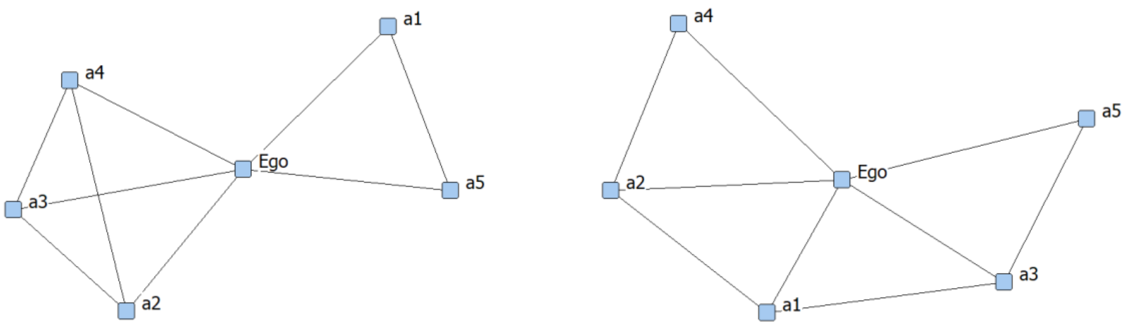


Density = 0.2 | Degrees = {2,1,1,0,0} | Constraint = 0.382
Term 1 = 0.2 | Term 2 = 0.16 | Term 3 = 0.04

Density = 0.2 | Degrees = {1,1,1,1,0} | Constraint = 0.4
Term 1 = 0.2 | Term 2 = 0.13 | Term 3 = 0.05

Figure 1. Two ego networks with identical densities but different constraint scores.

The third term is also a function of the degree sequence – indeed if we enumerate all ego networks with five alters, no graphs that have different degree sequences have the same score on the third term. However, in addition, the third term takes into account detailed information about the actual structure of the network, meaning that to calculate it we need to know the neighbourhoods of each of the alters. Figure 2 shows two ego networks with identical degree sequences, but different configurations, and therefore different term 3 scores. Consistent with standard brokerage imagery, the ego on the left is slightly less constrained than the one on the right. It is worth noting that ego network betweenness (Everett and Borgatti, 2005), also distinguishes between these two networks, assigning the left ego 12 points and the right ego just 9.



Density = 0.2 | Degrees = {2,2,2,1,1} | Constraint = 0.51
Term 1 = 0.2 | Term 2 = 0.24 | Term 3 = 0.07

Density = 0.2 | Degrees {2,2,2,1,1} | Constraint = 0.52
Term 1 = 0.2 | Term 2 = 0.24 | Term 3 = 0.08

Figure 2. Two ego networks with identical degree sequences but different Term 3 scores.

One benefit of the expression in equation (3) is that it allows us to see the relative magnitude of the terms. The first term has order 1/N, the ratio in the second term (provided it exists) is between 0.5 and 1 and so the order of this term is also 1/N. The order of the third term is more difficult to determine. If we look at the case in which one alter is connected to every other alter and no other alter-alter ties exist, then for the highly connected alter the last term would be $(N-1)^2/4N^2$ and this term would have a constant order. However, in many cases the terms inside the bracket would not be of order N so that this term would have order $1/N^2$. For example, if we enumerate all possible undirected ego networks with 6 alters, we find that when Term 2 is 0.185185, Term 3 can take on one of four possible values, depending on the structure of the ego network: 0.058642, 0.061728, 0.067901, and 0.177469. The last one is almost as large as Term 2.

The third term reaches its maximum value when the ego network contains a “shadow ego”, which occurs when one alter is connected to all others, and there are no other alter-alter ties as shown in Figure 3 (see also Burt, 2010, who refers to such networks as “partner networks”). In such a network, ego and the shadow ego are structurally equivalent, and all the other alters are structurally equivalent to each other.

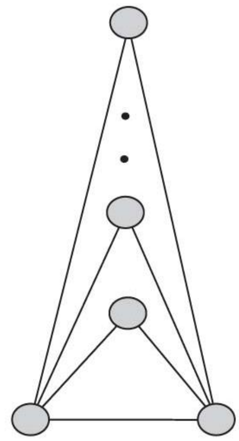


Figure 3 A shadow ego network

In this case, the maximum score for Term 3 is

$$\frac{(N-1)^2}{4N^2} + \frac{N-1}{N^4}$$

The maximum value for term 2 is easy to derive since increasing the degree of any alter increases the value of term 2 and so the maximum occurs when the deleted ego network is complete and is given by $2(N-1)/N^2$. As N increases then clearly it is possible for term 3 to dominate as term 2 must decrease.

The expression also allows us to easily see the minimum value for the constraint measure in undirected binary data. The minimum occurs when there are no alter-alter ties. Terms 2 and 3 vanish, and the constraint score is 1/N.

It is more difficult to derive the maximum value. When N is small, term 2 is able to dominate and so it would seem that a complete ego network would achieve the maximum. When N is large then term 3 can dominate and so a shadow ego network would achieve the maximum. Note the shadow ego network in the ego-deleted network would be a star. In Table 1 we give the constraint scores for 2 to 10 alters for both these networks.

Number of alters	Complete	Shadow
2	1.125	1.125
3	0.926	0.840
4	0.766	0.684
5	0.648	0.590
6	0.560	0.529
7	0.493	0.486
8	0.439	0.455
9	0.396	0.431
10	0.361	0.411

Table 1. Total constraint for complete and shadow ego networks

We can see that when N=8 the shadow ego network has higher constraint. We therefore conjecture that maximum constraint for N < 8 occurs in a complete network in which case equation (3) becomes

$$\frac{1}{N}+\frac{2}{N^2}\sum_{j=1}^{N}\frac{N-1}{N}+\frac{1}{N^2}\sum_{j=1}^{N}\left(\sum_{j=1}^{N-1}\frac{1}{N}\right)^2$$

$$=\frac{(2N-1)^2}{N^3}$$

If N is 8 or more than the maximum occurs in a shadow ego network in which case equation 3 becomes

$$\frac{1}{N}+\frac{2}{N^2}\left(\sum_{j=1}^{N-1}\frac{1}{2}+\frac{N-1}{N}\right)+\frac{1}{N^2}\left(\sum_{j=1}^{N-1}\frac{1}{N^2}+\left(\sum_{j=1}^{N-1}\frac{1}{2}\right)^2\right)$$

$$=\frac{(N+1)^2(N^2+4N-4)}{4N^4}$$

So to summarise we conjecture that maximum constraint is given by

$$\frac{(2N-1)^2}{N^3} \qquad 1 < N < 8$$

$$\frac{(N+1)^2(N^2+4N-4)}{4N^4} \qquad N \geq 8 \qquad (4)$$

As can be seen the maximum can be above 1, but only when N=2 in which case the maximum is 9/8. Burt (1992) indicates this in several places, including his Figure 2.3, which clearly shows constraint higher than 1 for the case of two alters and maximum density.

If we examine equation 4 we see that as N increases, the maximum value decreases; in fact as $N\rightarrow\infty$, maximum constraint tends to 0.25.

As noted constraint is at a maximum in small networks when the network is complete. This fact can be used to choose how to deal with isolates. Burt’s original definition does not apply to the case when ego is an isolate. However, applying Equation 3 to an isolate yields infinite constraint, which in practice could be set to just higher than the maximal possible (e.g., 1.2). This makes some sense as an isolate has no ties and therefore no social capital, which structural holes are meant to measure. Moreover, an isolate's neighbourhood is vacuously complete and so should have maximum constraint. However, if take the concept of constraint at face value – the condition of being constrained by the norms of those around you – then it makes sense to set constraint to zero as the isolate has complete autonomy of action. Thus, ultimately it depends on how constraint is being used theoretically.

We can also use the maximum and minimum values to normalize the measure given network size. If c denotes the raw constraint measure then a normalized measure c^* is given by

$$c^* = 1 \qquad N=0,1$$

$$c^* = \frac{N^2(Nc-1)}{3N^2-4N+1} \qquad 8>N>1$$

$$C^* = \frac{4N^3(Nc-1)}{(N-1)(N^3+3N^2+8N+4)} \qquad N \geq 8 \tag{5}$$

The c* measure is guaranteed to run between 0 and 1. This is appealing, but it should be remembered that this effectively removes network size from the measure of structural holes, which is inconsistent with its role as a measure of social capital.

.

4. Directed Binary and Valued Data

The extension to directed binary data is in general straight forward but a number of different cases need to be considered. The construction of P (see the formula just after Equation 1) is such that if A is a directed network then transposing A results in the same C. In fact reversing the direction of any arc in A will result in the same C provided the same alters are all still included in the ego network. It follows that in a directed network with no reciprocated ties we can simply ignore the direction and apply formula (3) on the underlying graph. For a network with all reciprocated ties we could again ignore the directions and use the underlying graph and we would get the same result.

Data with a mixture of reciprocated and unreciprocated ties needs to take full account of the reciprocity of each tie. Given an ego i and an unreciprocated alter j then $p_{ij}=1/\rho(i)$ where $\rho(i)$ is the total degree of i ie in-degree plus out-degree. For a reciprocated tie $p_{ij}=2/\rho(i)$. It follows that if we were to follow a similar argument as in the undirected case we would need to separate out the two cases for ego-alter ties into reciprocated and non-reciprocated. We would have to do the same with the alter-alter ties resulting in four different cases. If we have this information, then we can deconstruct constraint in a way similar to the undirected case. However, to achieve this we would require more information than in the undirected case. Instead of just degree for the first term, we would need both degree and the number of reciprocated ties. Similarly, for the second term we would need not only the degree sequences but also the number of reciprocated ties. The resulting formula -- although more complicated -- would still have the basic structure as the undirected case and the orders of the terms would still be the same. That is the third term would have a constant order and so would dominate as N increased for certain networks.

The minimum possible value of constraint would remain the same as the undirected case ie 1/N, the maximum however would change. In order to explore the maximum we examine both the complete and shadow ego networks which have a mixture of reciprocated and unreciprocated ties. If all ties were either reciprocated or all unreciprocated we would have exactly the same case as an undirected network. We therefore look for structures which have a higher constraint than the undirected case. In a complete network, as all the alters are equivalent then it is highly likely that all the ego-alter ties must either all be reciprocated or all be unreciprocated in order to make p_{ij} as large as possible ie 1/N. To maximize the alter –alter score all alters need to be connected to each other by a reciprocated edge. It follows that the maximum complete will most likely occur when the ego-alter ties are all not reciprocated.

In this case we can use this information to derive the maximum, which is given by equation (6). We note this gives higher values than in the undirected case.

$$\frac{1}{N}\left(\frac{4N-3}{2N-1}\right)^2 \quad (6)$$

As in the undirected case, the maximum occurs when N=2 giving a value of 1.388. In addition, there is an additional case that yields a value greater than 1, which is when N=3 the formula gives a maximum constraint of 1.08.

We now consider the shadow ego networks. We now have three types of nodes, ego, shadow ego and alters. This gives three types of edges. Ego-alter, ego-shadow ego, shadow ego-alter. Each of these edges can either be unreciprocated or reciprocated, giving rise to 6 different structures that have a mixture of ties. We experimentally tried the 6 different structures to see which had the highest constraint. This occurred when the ego-alter ties were unreciprocated and the ego-shadow ego tie and the shadow ego-alter ties were all reciprocated. An example of such a network is shown in Figure 4. It should be noted although the diagram shows the direction of the tie from ego to the alters any number of these could be reversed and the result would be the same.

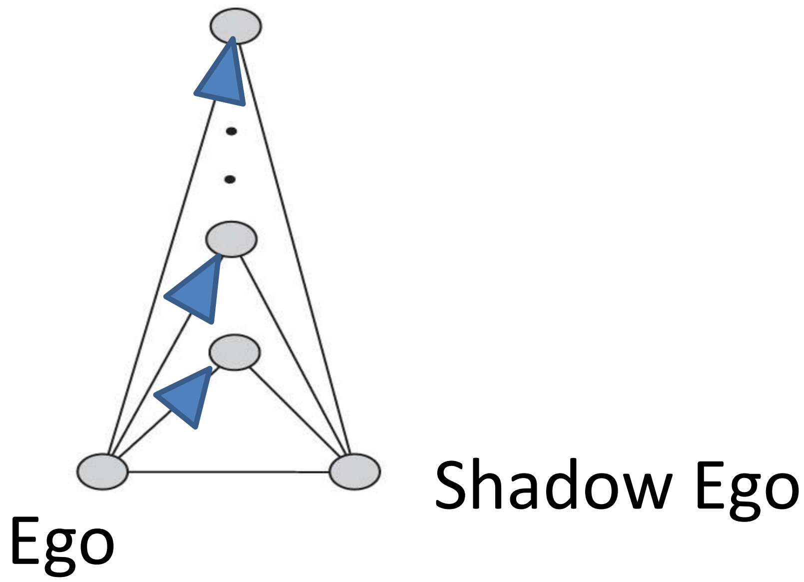


Figure 4. Maximum directed shadow ego network, note lines with no arrows are reciprocated

In this case, constraint is given by

$$\left(\frac{N+2}{3N(N+1)}\right)^2 (4N-3)(N+3)$$

As in the undirected case, we can see how both of these behave for networks of various sizes and this is given in Table 2

Number of alters	Complete	Shadow
2	1.388889	1.234568
3	1.08	1.041667
4	0.862245	0.91
5	0.71358	0.822716
6	0.607438	0.761905
7	0.528318	0.717474
8	0.467222	0.683728
9	0.418685	0.657284
10	0.379224	0.636033

Table 2 maximum constraint for complete and shadow ego networks

From Table 2 we see that the complete graph provides the maximum just for N=2 and N=3, whereas for N=4 or more, the shadow network has the maximum value. In fact, as N increases, constraint for the shadow ego network approaches 4/9 from above.

As in the undirected case we can use these values to normalize the measure as we did in equation 5..

We note that the complication in the directed case is due to the fact that, given the way the P matrix is constructed, the edges are in essence valued. We could for example construct a symmetric valued network by symmetrising the directed network using the sum. That is we would form $A+A^T$ so that reciprocated ties have a value of 2 and unreciprocated ties have a value of 1. Calculating all structural hole measures on this symmetric data would yield exactly the same results as the directed data.

When considering valued data (provided there is more than one value) then we need to understand that structural hole measures behave in a similar way to the directed case. This allows us to suggest a structure in which constraint is maximized. As seen from our directed data, constraint is at a maximum in a complete network when all ego alter ties are unreciprocated and all other ties are reciprocated. Rewording this in terms of valued data, we require that ego be connected to all alters with an unreciprocated tie of minimum value and that all the alters are connected by reciprocal ties of maximum value. This would make sense as we expect ego to be most constrained when they have weak connections to their alters all of whom have a strong connection to each other.

Now, for any directed ego network, suppose the smallest value is *m* and the largest value is M and there are N alters. Then the maximum value for complete network constraint is given by:

$$\frac{1}{N}\left(\frac{m+4M(N-1)}{m+2M(N-1)}\right)^2 \qquad (8)$$

We can see that, as M dominates, this term tends to 4/N, and as we require at least two alters the upper bound for complete constraint occurs when N=2 and is 2.

We apply the same reasoning for the shadow ego network. That is, ego is connected to all the alters (with the exception of shadow ego) with an unreciprocated tie with a value of *m* and all other ties are reciprocated with a value of M. In this case shadow constraint is given by

$$(N-1)\left(\frac{Nm+2M}{N(Nm-m+2M)}\right)^2+\left(\frac{2M(2M+Nm)}{(m+2M)(Nm-m+2M)}\right)^2$$

In this case, as M dominates, this term tends to $1+(N-1)/N^2$. We see that when N=2 or 3 the complete network can give higher maximum values (2 and 1.333) but for N=4 or more the shadow ego network has a higher maximum.

The minimum for valued data is still given by 1/(N). These values can again be used to normalize the measure.

For the undirected valued case the expression for the max is very similar but is now for the complete case

$$\frac{1}{N}\left(\frac{m+2M(N-1)}{m+M(N-1)}\right)^2 \quad (9)$$

And for the shadow case

$$(N-1)\left(\frac{Nm+M}{N(Nm-m+M)}\right)^2+\left(\frac{M(M+Nm)}{(m+M)(Nm-m+M)}\right)^2$$

and both have the same bounds as M increases as the directed case. The minimum is also the same as the directed case.

In the undirected case if m and M are close to each other then the complete network can give the maximum value provided N is less than 8. At which point the shadow ego network gives higher values than the complete network depends on how close *m* is to M. The same is true for the directed version, except now we require N to be less than 4.

5 Approximating Constraint for Undirected Binary data.

Suppose we are in the situation in which we know the degrees of the alters but we do not know the exact links. In fact it was precisely this situation that led the authors to look at this issue in detail. We were asked to advise someone who used a network self-assessment questionnaire to collect data as part of a self-reflexion exercise. In this exercise, respondents listed their contacts, then filled out a matrix of ties among the alters. What was turned in to the researcher/coach was just the number of contacts and the density. The advantage of such an instrument is that the respondent, doesn't need to reveal who they think is connected to whom. (It should be noted that asking respondents to simply estimate the density (ersatz network density) of their network has been shown to be unreliable Burt (1987)). After the fact, the researcher/coach wondered whether it would be possible to reconstruct constraint. The answer, of course, is no, as it is clear in our formulation that the actual alter alter ties are required. But let us first suppose that we had asked the respondents to turn in the degrees of the alters rather than the density. Confidentiality would still be preserved but we are now able to calculate the second term exactly. We already know that two ego networks with the same degree sequence can have different scores on the third term, so we know we cannot recover constraint precisely from this data alone. However, what if we can estimate Term 3 to reasonable degree? We will consider just the undirected case but clearly the same approach can be used for directed data. For the third term we do not know N(j) so cannot determine the q's. We can approximate ρ(q) by the average degree ⟨k⟩ which is equal to N-e, where e is Burt's effective size measure. In this case we have an approximation to the third term given by

$$\frac{1}{N^2}\sum_{j=1}^{N}\left(\frac{\rho(j)}{\langle k\rangle + 1}\right)^2$$

$$= \frac{1}{(N(\langle k\rangle + 1))^2}\sum_{j=1}^{N}\rho(j)^2 \qquad (10)$$

We shall call the approximated constraint measure given by replacing the third term in equation (3) by equation (10) approximation 1.

Suppose further, as in our given problem, that we only have the density and number of alters of the ego network. In this case clearly we know N and ⟨k⟩. We now replace ρ(j) by the average degree ⟨k⟩. In this case equation 3 reduces to

$$\frac{1}{N}\left(\frac{2\langle k\rangle + 1}{\langle k\rangle + 1}\right)^2 \qquad (11)$$

We shall refer to equation 11 as approximation 2. Given the fact that the third term is in many cases smaller than the first two terms we expect approximation 1 to be good in many cases.

To see how the approximations perform, Table 3 has the real constraint and the two approximations for the UCINET Think graph (Borgatti et al 2002) shown in Figure 5. Table 4 repeats the analysis for the UCINET Taro dataset Schwimmer E. (1973).

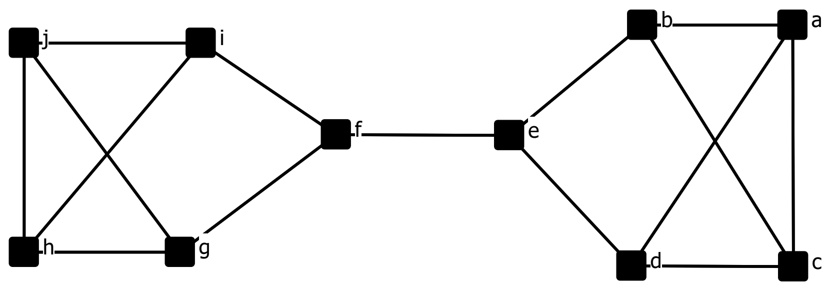


Figure 5 UCINET Think graph.

	Constraint	Approx 1	Approx 2
a	0.840	0.826	0.823
b	0.611	0.612	0.653
c	0.840	0.826	0.823
d	0.611	0.612	0.653
e	0.333	0.333	0.333
f	0.333	0.333	0.333
g	0.611	0.612	0.653
h	0.840	0.826	0.823
i	0.611	0.612	0.653
j	0.840	0.826	0.823

Table 3 Constraint with the two approximations for the Think graph

	Constraint	Approx 1	Approx 2
1	0.61	0.64	0.65
2	0.84	0.83	0.82
3	0.61	0.64	0.65
4	0.53	0.55	0.56
5	0.40	0.41	0.42
6	0.84	0.83	0.82
7	0.30	0.32	0.33
8	0.61	0.64	0.65
9	0.61	0.64	0.65
10	0.61	0.64	0.65
11	0.30	0.32	0.33
12	0.38	0.41	0.42
13	0.61	0.64	0.65
14	0.84	0.83	0.82
15	0.61	0.64	0.65
16	0.33	0.33	0.33
17	0.36	0.37	0.38
18	0.61	0.64	0.65
19	0.33	0.33	0.33
20	0.84	0.83	0.82
21	0.84	0.83	0.82
22	0.61	0.64	0.65

Table 4 Constraint and approximate constraint for the Taro data

We can see that in both cases both approximations are very good and in fact the correlations are all over 0.99. It is surprising that approximation 2 does so well given that it does not rely on structure at all.

However, we can construct networks in which the approximation is not good. In shadow ego graphs the third term dominates and this is not well captured by the approximation. As an extreme example consider a shadow ego network with 100 alters. In this case, one of the alters is connected to all other actors but no other alters are connected to each other. In this example the actual constraint is 0.265 but approximation 1 yields a value of 0.13 and approximation 2 gives 0.028. (See Table 5 for other values of N).

N	Real	Approx 1	Approx 2
5	0.59	0.54	0.52
10	0.41	0.32	0.27
20	0.33	0.22	0.14
30	0.30	0.18	0.09
40	0.29	0.16	0.07
50	0.28	0.15	0.06
60	0.28	0.15	0.05
70	0.27	0.14	0.04
80	0.27	0.14	0.03
90	0.27	0.13	0.03
100	0.27	0.13	0.03

Table 5 Constraint and approximate constraint for a shadow ego network

We do however have enough information to alert us that the approximation could be poor. The data we have for approximation 2 includes the size and density of the network. As we have seen, the shadow ego-deleted graph is a star and so is large and sparse. If we have the data as in approximation 1 then we have all the degree information. We could then calculate the degree centralization of the edge deleted graph and use this to weight a correction term. Approximation 1 is exact for a complete ego deleted graph and so we can add a correction term that would make it exact for an ego shadow graph. If we let C_D be the degree centralization of the ego deleted graph then the correction term is given by equation 12.

$$(N-1)\left(\frac{(N^3-N^2+4)}{4N^4}-\frac{N}{(3N-2)^2}\right)C_D \qquad (12)$$

If we have a complete graph then C_D is zero and so the correction term does not contribute to the approximation. For a shadow ego graph C_D will be 1 and so the approximation will be exact. However we recall that for N<7 the shadow graph must have a lower constraint than a complete graph, hence we only apply the correction for larger values say for N >10 As an example of the correction we look at the Kapferer Tailor shop data (Kapferer 1972) socializing ties at time period one. Then Abraham has 13 alters and the exact 3rd term is 0.08 the approximation is 0.07 and becomes 0.08 when we add the correction term.

If we had directed data, then deriving approximations similar to those in equation (11) would require information on the number of reciprocated ties. If such information were available it would seem likely that we would have access to the full data and so such approximations would be of very limited value. We therefore do not explore that further here.

6 Re-weighting

As highlighted in section 2, the third term in equation (3) is of a different order to the other terms. Although it often is of $O(1/N^2)$ and so is insignificant, when we have shadow ego networks it is of constant order. It could be argued that we should not combine the three aspects of the measure as they are capturing different properties. In some case we may have prior knowledge that tells us that we wish to weight one aspect more than another. In general this would be an unlikely scenario and it would be difficult to decide how much weight to give each of the terms.

Another possibility, instead of trying to choose the relative weighting of each term in advance, why not let a regression choose them for us? In other words, regress an outcome variable such as performance on all three terms. The regression would then combine them optimally to predict performance. This will result in a better r-square (or, at least, never worse), and give us diagnostic information about which term is important in a given research setting. Using data from Parker, Halgin and Borgatti (2016), we regressed employee performance ratings both on constraint (results in Table 6), and on the elements of constraint separately. The improvement in r-squared is 33% (although the numbers are so small this could be just noise). It can be seen that, in this dataset, the effect of Term 3 is negligible, and that Terms 1 and 2 both contribute. In other datasets we have seen, only Term 1 was related to the outcome variable

Table 6. Regressing performance on constraint and the elements of constraint. N = 554

Model 1	Coef.	Std. Err.	t	P>t
Constraint	-10.6580	1.9234	-5.54	0.000
Intercept	0.4304	0.0880	4.89	0.000
R-squared =	0.0527			

Model 2	Coef.	Std. Err.	t	P>t
term1	3234.77	1074.52	3.01	0.003
term2	-1656.59	542.75	-3.05	0.002
term3	-44.03	158.61	-0.28	0.781
Intercept	0.78	0.17	4.63	0.000
R-squared =	0.0701			

Unfortunately, the three terms are far from orthogonal, creating serious multi-collinearity problems. This is especially true of terms 2 and 3, which both depend heavily on the degree distribution.

7 Discussion

In this work, we have sought to reformulate Burt’s constraint measure in order to provide a better appreciation of what exactly it measures. One result of this effort was the discovery that, for larger ego networks, the popular image of constraint as a measure of ego net closure is incorrect. It has been widely assumed in the literature that the maximum constraint score occurs in the case of a maximally dense/closed ego network (in which every alter is connected to every other). However, in reality, even a relatively sparse ego network can have a higher constraint score if it is the right shape. For ego networks with more than 7 alters, the highest possible constraint score occurs when there is one alter that is connected to every other alter, and when there are no other alter-alter ties. We refer to the alter that is connected to the other alters as a "shadow ego". A shadow ego is effectively a second broker that brokers the exact same alters that ego does – indeed, they are structurally equivalent within the ego network.

The constraint measure asserts (for $N > 7$) that an ego network with a shadow-ego shape has less social capital than a similar-sized network in which all alters are connected. Whether this makes sense empirically is an interesting thing. Burt has pointed out that, although highly constraining, shadow ego networks (which he refers to as "partner networks" or "borrowed entrepreneurial networks") can be advantageous for women. Burt argues that women suffer a legitimacy deficit in the workplace, with the result that having a sponsor who vouches for them and lends them their own social capital is helpful, in spite creating a maximum constraint score.

We also note that shadow ego networks exhibit another property that is perhaps counter intuitive. By definition, networks with large effective size are seen as having a lot of structural holes and hence low constraint. But, an undirected shadow ego network has an effective size of $N-2(N-1)/N$ which is very close to N, where N is the maximum possible. Yet this network also has maximal constraint, which is an inverse measure of structural holes. This fact demonstrates why it is important to consider both of Burt's measures. The key, of course, is that effective size does not take account of hierarchy, whereas constraint does. Which one to use will depend on the theoretical mechanisms proposed in a given study.

Finally, an interesting aspect of constraint that is revealed by shadow ego networks is its extreme sensitivity to the presence or absence of a single tie. Suppose we have a shadow ego network with N, the number of alters, sufficiently large that constraint achieves its maximum score. Deleting the edge that connects ego to the shadow ego would result in a network with minimum constraint. Unlike effective size, constraint is highly sensitive to the deletion of (or failure to measure) a single edge. It is counter-intuitive that ego breaking a tie with a shadow ego would remove all the constraint that the shadow ego has imparted because whether or not ego is connected to the shadow ego, the shadow ego is still there, still brokering the same alters, and therefore still reducing advantage for ego. This is unlike the behaviour of 2-step betweenness (Borgatti and Everett, 2006; Brandes, 2008), which is ordinary betweenness restricted to geodesics of length 2. Loosely speaking, the measure looks at every pair of nodes that is separated by just one intermediary, and calculates to what extent ego is an exclusive intermediary, as opposed to just one of many. For example, consider the ego shadow network in Figure 6.

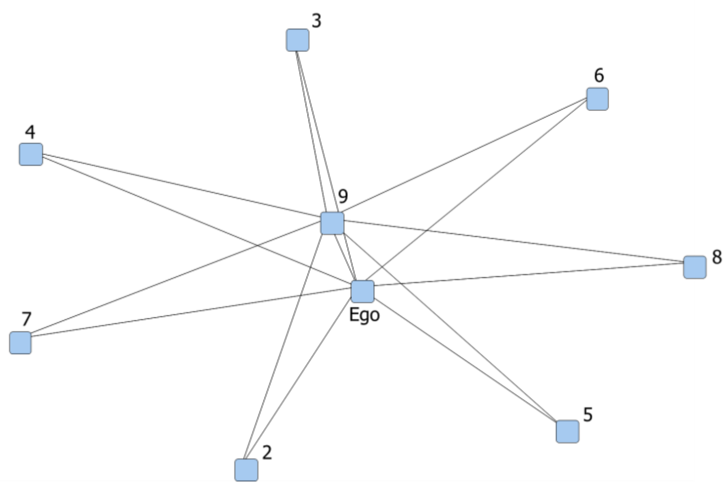


Figure 6. A shadow ego network

In the network, there are two equally short paths from 6 to 7. One of these passes through ego, so ego gets a half a point for this. Summing across all 21 pairs, this yields a 2-step betweenness score of 21. Constraint is 0.455, the highest possible. Now suppose we delete the tie from ego to 9, so that 9 is no longer a part of ego's network. Constraint now plummets to .143, the minimum possible. However, 2-step betweenness is unchanged at 21. Unlike constraint, the 2-step betweenness measure captures the reality that the other broker reduces ego's advantage regardless of whether ego is connected to that person.

References

Borgatti, S. P. 1997. Structural holes: Unpacking Burt's redundancy measures. *Connections*, 20(1): 35-38

Borgatti, S. P., Everett, M. G., 2006. A graph-theoretic perspective on centrality. *Social Networks* 28, 466–484

Borgatti, S. P., Everett, M. G. and Freeman, L. C. (2002) 'Ucinet for Windows: Software for Social Network Analysis.' Harvard: MA:Analytic Technologies.

Brandes, U., 2008. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2), pp.136-145.

Burt, R.S., 1987. A note on the general social survey's ersatz network density item. *Social Networks*, *9*(1), pp.75-85.

Burt, R.S. 1992. *Structural Holes: The social structure of competition.* Cambridge: Harvard University Press.

Burt, R. S. 1998. The gender of social capital. *Rationality and society*, *10*(1), 5-46.

Burt, R.S., 2010. Neighbor networks: Competitive advantage local and personal. Oxford University Press.

Everett, M. and Borgatti, S.P., 2005. Ego network betweenness. *Social networks*, *27*(1), pp.31-38.

Kapferer B. (1972). Strategy and transaction in an African factory. Manchester: Manchester University Press.

Parker, A., Halgin, D. S., & Borgatti, S. P. (2016). Dynamics of social capital: Effects of performance feedback on network change. *Organization Studies*, *37*(3), 375-397.

Schwimmer E. (1973). Exchange in the social structure of the Orokaiva. New York: St Martins.