Calculating RSS for Finite Discovery Ranges

Stephen R. Haptonstahl

February 25, 2014

This implements the Relation Strength Similarity (RSS) measure defined in Chen et al. (2012).

1 Assumptions

Let B be the adjacency matrix of the input network where b_{ij} represents the strength of the link from node i to node j. First, divide each row by its row sum, so the sum of each row is now 1. This normalization implements Equation (1) of Chen et al. (2002). Call this normalized adjacency matrix A.

Now for each directed dyad (i, j) we calculate the RSS score $S(i, j)$ as

$$S(i,j) = \sum_{k=0}^{r} L(i,j,k)$$

where $r \geq 1$ is the discovery range (the maximum path length considered between i and j) and $L(i, j, k)$ is the product of $a_{\cdot,\cdot}$ along simple[1] paths of length k from i to j.

1.1 $r = 0$

We will not search paths of length zero (from ii to i) but it may be useful to define $S(i, i)$.

$$L(i,j,0) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

This serves two purposes: (1) It defines $S(i, i) = 1$ which fits our notion that RSS measures strength of connection between two nodes, and (2) for $k > 0$ it defines $L(i, i, k) = 0$ which is correct (every path from ii to i repeats a node, manely i, so it's not simple and it therefore excluded) and convenient notationally.

1.2 $r = 1$

The only path (simple or otherwise) between nodes i and j of length 1 is given by the link between them.

$$L(i,j,1) = a_{ij} \tag{2}$$

[1] A simple path is one that does not contain any node twice.

1.3 $r = 2$

We now consider paths of length 2 that have some node ℓ between nodes i and j. Considering only simple paths means here only that $\ell \neq i$ and $\ell \neq j$.

$$
\begin{aligned}
L(i,j,2) &= \sum_{\ell \neq i,j} a_{i\ell} a_{\ell j} && (3)\\
&= \sum_{\ell=1}^{n} a_{i\ell} a_{\ell j} - a_{ii} a_{ij} - a_{ij} a_{jj} && (4)\\
&= \sum_{\ell=1}^{n} a_{i\ell} a_{\ell j} && (5)
\end{aligned}
$$

if we set all diagonal entries to zero. This makes sense given that we are only considering simple paths, so this should be part of the initial normalization of the adjacency matrix. However, remember that $S(i,i) = 1$, not zero, for all i, and this case will have to be trapped separately from the calculations using $L(\cdot)$.

1.4 $r = 3$

We now consider paths of length 3 that follow nodes $i \to m \to \ell \to j$. Considering only simple paths means here only that $\ell \neq i,j$ and $m \neq i,j,\ell$.

$$
\begin{aligned}
L(i,j,3) &= \sum_{\ell \neq i,j} \sum_{m \neq i,j,\ell} a_{im} a_{m\ell} a_{\ell j} && (6)\\
&= \sum_{\ell \neq i,j} a_{\ell j} \left[\sum_{m \neq i,j,\ell} a_{im} a_{m\ell} \right] && (7)\\
&= \sum_{\ell \neq i,j} a_{\ell j} \left[\sum_{m=1}^{n} a_{im} a_{m\ell} - a_{ij} a_{j\ell} \right] && (8)\\
&= \sum_{\ell \neq i,j} a_{\ell j} \left[\text{Ł}(i,\ell,2) - a_{ij} a_{j\ell} \right] && (9)\\
&= \sum_{\ell=1}^{n} a_{\ell j} \left[\text{Ł}(i,\ell,2) - a_{ij} a_{j\ell} \right] && (10)\\
&= \text{sum} \left(A[,j] \cdot \begin{bmatrix} L(i,1,2) - a_{ij} a_{j1} \\ L(i,2,2) - a_{ij} a_{j2} \\ \vdots \\ L(i,n,2) - a_{ij} a_{jn} \end{bmatrix} \right) - a_{ij} [\underbrace{L(i,i,2)}_{=0} - a_{ij} a_{ji}] && (11)\\
&= \text{sum} \left(A[,j] \cdot \begin{bmatrix} L(i,1,2) - a_{ij} a_{j1} \\ L(i,2,2) - a_{ij} a_{j2} \\ \vdots \\ L(i,n,2) - a_{ij} a_{jn} \end{bmatrix} \right) + a_{ij}^2 a_{ji} && (12)
\end{aligned}
$$

1.5 $r = 4$

We now consider paths of length 4 that follow nodes $i \to p \to m \to \ell \to j$. Considering only simple paths means here only that $\ell \neq i, j$, $m \neq i, j, \ell$, and $p \neq i, j, \ell, m$.

$$\begin{aligned}
L(i,j,4) &= \sum_{\ell \neq i,j} \sum_{m \neq i,j,\ell} \sum_{p \neq i,j,\ell,m} a_{ip} a_{pm} a_{m\ell} a_{\ell j} && (13)\\
&= \sum_{\ell \neq i,j} a_{\ell j} \left[\sum_{m \neq i,j,\ell} \sum_{p \neq i,j,\ell,m} a_{ip} a_{pm} a_{m\ell} \right] && (14)\\
&= \sum_{\ell \neq i,j} a_{\ell j} \left[\sum_{m \neq i,j,\ell} \left(\sum_{p \neq i,\ell,m} a_{ip} a_{pm} a_{m\ell} - a_{ij} a_{jm} a_{m\ell} \right) \right] && (15)\\
&= \sum_{\ell \neq i,j} a_{\ell j} \left[\sum_{m \neq i,j,\ell} \sum_{p \neq i,\ell,m} a_{ip} a_{pm} a_{m\ell} - \sum_{m \neq i,j,\ell} a_{ij} a_{jm} a_{m\ell} \right] && (16)\\
&= \sum_{\ell \neq i,j} a_{\ell j} \left[\sum_{m \neq i,\ell} \sum_{p \neq i,\ell,m} a_{ip} a_{pm} a_{m\ell} - \sum_{p \neq i,j,\ell} a_{ip} a_{pj} a_{j\ell} - \sum_{m \neq i,j,\ell} a_{ij} a_{jm} a_{m\ell} \right] && (17)\\
&= \sum_{\ell \neq i,j} a_{\ell j} \left[L(i,\ell,3) - a_{j\ell} \sum_{p \neq i,j,\ell} a_{ip} a_{pj} - a_{ij} \sum_{m \neq i,j,\ell} a_{jm} a_{m\ell} \right] && (18)\\
&= \sum_{\ell \neq i,j} a_{\ell j} \Bigg[L(i,\ell,3) - a_{j\ell} \Big(\underbrace{\mathrm{sum}(A[i,] \cdot A[,j])}_{L(i,j,2)} - a_{i\ell} a_{\ell j} \Big) \\
&\qquad - a_{ij} \Big(\underbrace{\mathrm{sum}(A[j,] \cdot A[,\ell])}_{L(j,\ell,2)} - a_{ji} a_{i\ell} \Big) \Bigg] && (19)\\
&= \sum_{\ell \neq i,j} a_{\ell j} \Big[L(i,\ell,3) - a_{j\ell} L(i,j,2) + a_{j\ell} a_{i\ell} a_{\ell j} - a_{ij} L(j,\ell,2) + a_{ij} a_{ji} a_{i\ell} \Big] && (20)\\
&= \sum_{\ell \neq i} a_{\ell j} \Big[L(i,\ell,3) - a_{j\ell} L(i,j,2) + a_{j\ell} a_{i\ell} a_{\ell j} - a_{ij} L(j,\ell,2) + a_{ij} a_{ji} a_{i\ell} \Big] && (21)
\end{aligned}$$

because $a_{jj} = 0$. When $\ell = i$ some terms are zero and some are not, so

$$\begin{aligned}
&= \sum_{\ell=1}^{n} a_{\ell j} \Big[L(i,\ell,3) - a_{j\ell} L(i,j,2) + a_{j\ell} a_{i\ell} a_{\ell j} - a_{ij} L(j,\ell,2) + a_{ij} a_{ji} a_{i\ell} \Big] \\
&\qquad + a_{ij} a_{ji} L(i,j,2) + a_{ij}^2 L(j,i,2) && (22)\\
&= \mathrm{sum} \left(A[,j] \cdot \begin{bmatrix} L(i,1,3) - a_{j1} L(i,j,2) - a_{ij} L(j,1,2) + a_{j1} a_{i1} a_{1j} + a_{ij} a_{ji} a_{i1} \\ L(i,2,3) - a_{j2} L(i,j,2) - a_{ij} L(j,2,2) + a_{j2} a_{i2} a_{2j} + a_{ij} a_{ji} a_{i2} \\ \vdots \\ L(i,n,3) - a_{jn} L(i,j,2) - a_{ij} L(j,n,2) + a_{jn} a_{in} a_{nj} + a_{ij} a_{ji} a_{in} \end{bmatrix} \right) \\
&\qquad + a_{ij} a_{ji} L(i,j,2) + a_{ij}^2 L(j,i,2) && (23)
\end{aligned}$$

1.6 $r = 5$

We now consider paths of length 5 that follow nodes $i \to q \to p \to m \to \ell \to j$. Considering only simple paths means here only that $\ell \neq i, j$, $m \neq i, j, \ell$, $p \neq i, j, \ell, m$, and $q \neq i, j, \ell, m, p$.

$$L(i, j, 5) = \sum_{\ell \neq i,j} \sum_{m \neq i,j,\ell} \sum_{p \neq i,j,\ell,m} \sum_{q \neq i,j,\ell,m,p} a_{iq} a_{qp} a_{pm} a_{m\ell} a_{\ell j} \tag{24}$$

$$\tag{25}$$