Social Networks are Encoded in Language

Sterling Hutchinson (schtchns@memphis.edu)
Department of Psychology / Institute for Intelligent Systems, University of Memphis
365 Innovation Drive, Memphis, TN 38152 USA

Vivek Datla (vvdatla@memphis.edu)
Department of Computer Science / Institute for Intelligent Systems, University of Memphis
365 Innovation Drive, Memphis, TN 38152 USA

Max M. Louwerse (mlouwerse@memphis.edu)
Department of Psychology / Institute for Intelligent Systems, University of Memphis
365 Innovation Drive, Memphis, TN 38152 USA

Abstract

Knowledge regarding social information is thought to be derived from many different sources, such as interviews and formal relationships. Social networks can likewise be generated from such external information. Recent work has demonstrated that statistical linguistic data can explain findings thought to be explained by external factors alone, such as perceptual relations. The current study explored whether language implicitly comprises information that allows for extracting social networks, by testing the hypothesis that individuals who are socially related together are linguistically talked about together, as well as the hypothesis that individuals who are socially related more are talked about more. In the first analysis using first-order co-occurrences of names of characters in the *Harry Potter* novels we found that an MDS solution correlated with the actual social network of characters as rated by humans. In a second study using higher-order co-occurrences, a latent semantic analysis (LSA) space was trained on all seven *Harry Potter* novels. LSA cosine values for all character pairs were obtained, marking their semantic similarity. Again, an MDS analysis comparing the LSA data with the actual social relationships yielded a significant bidimensional regression. These results demonstrate that linguistic information indeed encodes social relationship information and show that implicit information within language can generate social networks.

Keywords: social relations; social networks; social cognition; statistical linguistic frequencies

Introduction

What is the nature of social relations and how can such relations be estimated? Social media, such as Facebook, LinkedIn, and Twitter allow us to answer this question, based on individuals choosing their friends. However, when such deliberate decisions are not readily available, how can social relations be measured and social networks be plotted otherwise?

Social relations can be interpreted in three non-mutually exclusive ways (Fischer, 1982). First, they can be formal in socially recognized roles, such as teacher/student, employer/employee, or father/son. Second, they can be sentimental, as when individuals feel close to others. Finally, a relation can be defined in terms of interactions and exchanges. The formal, sentimental, and interactive nature of the social relationship can be determined by assessing a number of factors. For example, relationships can be predicted in part by the kinship of the individuals. In families, siblings tend to be close friends. Gender especially impacts the nature of relationships such that if a member of the dyad is a female, the relationship is more likely to be successful (Kim, McHale, & Osgood, 2006; Wright & Scanlon, 1991). Environment tends to weigh heavily in terms of whether or not two individuals are likely to build a relationship together. Proximity has also long been established as a strong predictor for relationships of all varieties, with increased proximity leading to increased likelihood of interpersonal relationships (Ebbesen, Kjos, Konecni, 1976). In addition, ties between locations (e.g., commonly trekked routes) also impact social interaction (Takhteyev, Gruzd, & Wellman, 2011). Similarly, familiarity fosters attraction between individuals (Reis, Manianci, Caprariello, Eastwick, & Finkel, 2011; Zajonc, 1968; 2001). Further, those who share interests, attitudes, and characteristics are more likely to develop friendships. In fact, any similarity between two individuals promotes the formation of a relationship between them (Bryne, 1971), with important matters (e.g., religious views, political attitudes) given more weight (Touhey, 1972). Emotions also impact relationships. When two individuals first encounter one another, a future friendship becomes more likely if the interaction is positive, whereas a friendship is not apt to blossom if the interaction is negative (Farina, Wheeler, & Mehta, 1991). Even physical features, like smell or appearance influence the relationships we form (Li, Moallem, Paller, & Gottfried, 2007).

After social relations are formed, different factors help these relations to solidify. For instance, Berscheid, Snyder, and Omoto (1989) found that closeness was significantly related to satisfaction of established romantic relationships, as was self-disclosure (Sprecher & Henrick, 2004). Feeney and Noller (1992) argued that individual differences like attachment styles impact the duration of social relationships, as does equity (Hatfield, Traupmann, & Walster, 1978). In addition, when it comes to group relationships, predicting

outcomes becomes more complicated, with members constantly joining and leaving the group. In fact, Kariam, Wang, and Leskovec (2012) found that diffusion growth (i.e., the addition of a new members to a group due to current relationships with group members) limits group size. Such findings indicate that the structure of groups is impacted by various factors.

When these social relationships are formed, how do we plot these relationships as social networks? Social scientists typically rely on interviews. For instance, Fischer (1982) asked respondents who they would share personal information with, who they worked with, and who visited their house, etc. in order to plot personal networks in the San Francisco area. But how could such networks be represented when participants cannot be interviewed as in Fischer's study, or when participants otherwise do not voluntarily release personal information as in social media? One answer to this question might lie in language. Perhaps social networks can also be acquired from and represented implicitly through linguistic sources.

In several studies we have demonstrated that perceptual information that is readily available from the world around us is encoded in language. For instance, Louwerse, Cai, Hu, Ventura, and Jeuniaux (2006) and Louwerse and Zwaan (2009) tested whether language encodes geographical information by correlating statistical linguistic frequencies between cities with the actual physical distances between those cities. Louwerse and Zwaan (2009) further tested the hypothesis by correlating computationally generated semantic relationship cosine values with the longitude and latitude of cities in the US. Recently, Louwerse, Hutchinson, and Cai (in press) found evidence for this hypothesis with Chinese texts predicting locations in China and Arabic texts predicting locations in the Middle East. Louwerse and Benesh (in press) have further extended these findings by demonstrating that the longitude and latitude of cities in the fictional Middle Earth can be predicted using the text of the *Lord of the Rings* trilogy. The semantic associations between cities in a corpus accurately estimated the physical distance between cities, supporting the claim that language encodes geographical information. A similar reasoning can perhaps be applied to social relationships. If the physical distance between individuals is small, their semantic association might be high.

In a number of studies we have shown that perceptual and embodied relations are encoded in language (Louwerse, 2011). Perhaps social relations are also encoded in language, such that computational algorithms can extract such relations from text. In the current paper we tested two hypotheses. First, if individuals are socially related together, they are talked about together. That is, if individuals are in social proximity, they are likely to be found in textual proximity. Second, if individuals are socially related more, they are talked about more.

To test these hypotheses we used the seven fantasy fiction *Harry Potter* novels and extracted the semantic relationships between the characters and compared them with the actual social relationships between the characters.

The Social Network of Harry Potter

Social networks are structures that map relationships between individuals. They are complex systems that can be used to examine, predict, and measure various features embedded within a network (see Newman, 2003 for an overview). Nodes represent specific individuals with edges connecting those individuals and representing relational information.

There are several ways social networks are produced. Social networks are often generated manually whereby individuals are linked to others if they are friends, colleagues, family members, etc. Individuals are able to generate their own egocentric social networks representing those other individuals with whom they share a relationship. Of course, the individual generating the network will do so based on the existence and strength of relationships that were generated by and subject to the factors enumerated above (Scott, 1988). This is the technique employed by Muckety LLC (2012). Muckety is a news corporation that manually generates maps of relationship influence between relevant individuals in a network. They manually specify networks of influence where each node is related to numerous other nodes via specific types of relationships (e.g., friend, enemy, relative). These relationships are manually researched using a variety of sources, such as government agencies and organizations, news publications, books, organization web sites, and interviews, and are expectedly costly to produce. Although Muckety generally generates networks representing current political, financial, and educational communities they have also constructed a social network representing each of the relationships between characters and organizations from all Harry Potter novels (Rowling, 1998; 1999a; 1999b; 2000; 2003; 2005; 2007).

Although Muckety provided a manually generated relationship network, edge weights between nodes were not provided. We thus computed edge weights as follows. Considering that between any two individuals there exists approximately four friendship links (Backstrom, Boldi, Rosa, Uander, & Vigna, 2012), we calculated a value representing higher-order relationships four degrees away. First order relationships were assigned a value of 1, relationships separated by one friendship link (or degree of separation) were assigned a value of .5, relationships separated by two friendship links were assigned a value of .25, relationships separated by three friendship links were assigned a value of .125, and relationships separated by four friendship links were assigned a value of .0625. To illustrate, *Harry Potter – Ron Weasley* received +1 because they are directly related as friends. *Harry Potter – Percy Weasley* received + .5 because they both share a relationship with *Ron Weasley*. *Harry Potter – Igor Karkaroff* would receive + .25 because *Harry* shares a relationship with someone (e.g., *Dumbledore*) who shares a relationship with

another person (e.g. *Snape*) who directly shares a relationship with *Igor Karkaroff.* This process was repeated until four friendship links were reached.

Computational Study

The current study investigated whether statistical linguistic information encodes social relationships by testing the hypotheses a) if individuals are socially related together, they talked about together, b) if individuals are socially related more, they are talked about more. Two computational algorithms were used to test these hypotheses. First, we relied on first-order co-occurrence frequencies of character names. Although first-order frequencies are easy to compute, they also come at a price. Due to sparsity problems, they can sometimes give a biased result (Louwerse, 2011). We therefore also used a higher-order co-occurrence algorithm, latent semantic analysis (LSA; Landauer, McNamara, Dennis & Kintsch, 2007).

The seven Harry Potter books were converted to one electronic document used for the research purposes described in this study only. The document consisted of a total of 1,277,991, words. The electronic document was then filtered and all stop words (grammatical items) and punctuation marks were removed, resulting in a final file with 517,501 words and 21,423 paragraphs.

First-order co-occurrences

In order to determine the first-order co-occurrences of character names, we computed the co-occurrence of all combinations of 56 character names in the Harry Potter novels with name pair in a five-word window. To avoid any biases with single word and two-word names (*Harry* versus *Harry Potter*), we selected the names by which each character was most frequently called while keeping the least ambiguous (e.g., *Ron Weasley* and *Arthur Weasley* are both referred to as *Weasley*, we therefore selected the names *Ron* and *Arthur).*

Although Muckety included 263 nodes (including characters, organizations, and locations), we were only interested in character relationships. We selected 56 characters for the analysis to keep the analysis from becoming too computationally expensive, as each character was paired with every other character. Characters were included on the basis of their prominence in the Harry Potter series, i.e., obscure characters were excluded from the analysis as they shared relationships with the fewest other characters.

These 56 x 56 frequency combinations were entered in an MDS analysis using the SMACOF algorithm. The SMACOF algorithm minimizes the sum of squares of the error by optimizing the fit to the distances (as opposed to the squared distances) and is thus preferred to ALSCAL (Young, 1985). We used default criteria for SMACOF, with the maximum iterations = 100, stress convergence = .0001 and the minimum stress value = .0001. Co-occurrence frequencies converged in 10 iterations with stress = .16.

Similarly, the Muckety scores for all 56 x 56 relations were entered in an MDS analysis, using the same parameters as for the linguistic data. The MDS converged in 25 iterations, with stress = .13.

To do justice to the 2D structure of the Muckety data, we conducted a bidimensional regression to determine the relationship between the human data and the statistical linguistic frequency data. Tobler (1964) and Friedman & Kohler (2003) introduced bidimensional regressions in order to compute the mapping of any two planes under consideration. Whereas in a unidimensional regression each data point is shifted by intercept and slope, each actual and predicted value of the dependent variable are presented by a point in space, whereby vectors represent intercept and slope.

The bidimensional regression for Muckety and co-occurrence values yielded a moderate correlation, $r = .43$, $p < .001$, $n = 56$. The moderate correlation can most likely be attributed to the relatively small size of the corpus, as this impacts first order co-occurrences most (Louwerse, 2011). To account for this sparsity problem, it is often recommend to not so much rely on first-order co-occurrences, but on higher-order co-occurrences.

Higher-order co-occurrences

To compute the higher-order computational relationship strength values we employed Latent semantic analysis (LSA). Latent Semantic Analysis captures semantic relations by mapping initially meaningless words into a continuous high dimensional semantic space (Landauer, McNamara, Dennis & Kintsch, 2007). More specifically, a first-order process associates stimuli (words) and the contexts they occur in (paragraphs). Stimuli are paired based on their contiguity or co-occurrence. These local associations are next transformed by means of Singular Value Decomposition (SVD) into a small number of dimensions (typically 300) yielding more unified knowledge representations by removing noise.

In the current study the input was the electronic version of the Harry Potter novels, segmented into paragraphs, from which a large term-document was created. For example, if there are m terms in n paragraphs, a matrix of $A = (f_{ij}\text{xG}(j)\text{xL}(i,j))_{mxn}$ is obtained. The value of f_{ij} is a function of the integer that represents the number of times term i appears in document j, $L(i; j)$ is a local weighting of term i in document j, and $G(j)$ is the global weighting for term j. Such a weighting function is used to differentially treat terms and documents to reflect knowledge that is beyond the collection of the documents. The large matrix of A has, however, much redundant information, for instance because not every word occurs in every paragraph. Singular Value Decomposition reduces this noise by decomposing the matrix A into three matrices $A = U\Sigma V'$; where U is an m by m square matrix and V is an n by n square matrix, with Σ being an m by n diagonal matrix with singular values on the diagonal. By removing dimensions corresponding to smaller singular values, the representation of each word is reduced as a smaller vector with each word now becomes a weighted

Figure 1: MDS loadings of Muckety scores (x-axis) and LSA higher-order values (y-axis) for the first dimension (Figure 1A) and second dimension (Figure 1B)

vector on 300 dimensions, with only the most important dimensions that correspond to larger singular values being kept (Landauer et al., 2007). The semantic relationship between characters can then be estimated by taking the cosine between two vectors.

As before, the 56 x 56 cosine matrix was submitted to MDS, which converged in 11 iterations with stress = .15. When we compared the two-dimensional loadings of the Muckety scores and the LSA scores in a bidimensional regression, we found, somewhat surprisingly, a weak correlation, r =.23, p = .08, n = 56. Yet when LSA values were allowed to populate a three dimensional configuration (stress = .07, convergence in 20 iterations), the bi-dimensional regression between Muckety scores and the second and third dimension of the LSA MDS yielded a more moderate (and significant) correlation, r = .30, p = .02, n = 56.

Upon visual inspection of the MDS plot, the first dimension did not explain social relations, but identified an outlier in the data. The character *Ginny Weasley* had more direct relationships than any of the other characters (except Harry Potter), yet the frequency with which *Ginny* occurred in the text was quite low. To illustrate, the word *Harry* occurred 21,781 times in the text whereas the word *Ginny* only occurred 762 times. After the removal of this outlier we again ran an MDS with two dimensions for both LSA (stress = .13, convergence in 12 iterations) and Muckety values (stress = .09, convergence in 33 iterations) (see Figure 1). The bidimensional regression now yielded a strong correlation between LSA values and Muckety values, r = .76, p < .001, n =55.

The Muckety standard and the LSA estimates are illustrated in Figures 1A and 1B. As both figures show, the correlation between Muckety and LSA loadings are relatively strong. The correlation between Muckety and LSA values for the first dimension is represented in Figure 1A, r = .56, and the correlation between Muckety and LSA values for the second dimension is represented in Figure 1B, r = .76.

Both the first-order and the higher-order co-occurrence results demonstrate that it is possible to extract a social network from language using statistical linguistic frequencies of names of individuals.

Number of Relationships

In this study we also test the hypothesis that characters who were socially related more are talked about more. We therefore computed the frequency of the 56 character used in the previous study. Next, we calculated how many relationships each character had in the Muckety network. For instance, *Harry Potter* shared direct relationships with 37 other characters whereas *Luna Lovegood* only shared direct relationships with 9 other characters (see Table 1).

Name frequency and number of relationships correlated strongly, r = .72, p < .001, n = 56, suggesting that individuals who have a large social network appear more in the text.

Discussion

The current study aimed to determine if language encodes social relationship information. The reported results suggest computationally derived character pair values can explain human relationship strength ratings. If linguistic representations did not encode relationship structures we would have expected our computationally derived relationship strength values not to predict the human scores. They do, however, significantly predict the human Muckety relationship network scores, suggesting that social information is encoded in language, supporting the hypotheses that individuals who are socially related together are talked about together, and those who are socially related more are talked about more.

Table 1: Number of Relationships (Rel) and Frequency (Freq) of Character Names

Character Name	Rel	Freq	Character Name	Rel	Freq	Character Name	Rel	Freq
Aberforth Dumbledore	6	78	Fred Weasley	16	1075	Narcissa Malfoy	8	75
Alastor Moody	6	874	George Weasley	16	898	Neville Longbottom	10	928
Albus Dumbledore	16	3981	Ginny Weasley	19	792	Nymphadora Tonks	9	243
Amos Diggory	4	54	Goyle	3	278	Padma Patil	5	35
Argus Filch	3	335	Harry Potter	37	21781	Parvati Patil	5	168
Arthur Weasley	15	171	Hermione Granger	13	6132	Percy Weasley	7	512
Augusta Longbottom	3	2	Igor Karkaroff	3	321	Peter Pettigrew	8	163
Bellatrix Lestrange	16	250	James Potter	10	186	Petunia Dursley	5	671
Bill Weasley	13	365	Kingsley Shacklebolt	4	119	Remus Lupin	12	841
Buckbeak	3	131	Kreacher	3	305	Ron Weasley	17	9144
Cedric Diggory	9	813	Lavender Brown	5	286	Rubeus Hagrid	12	2342
Charlie Weasley	12	165	Lily Potter	10	119	Severus Snape	15	2172
Cho Chang	9	261	Lucius Malfoy	10	148	Sirius Black	15	2314
Cornelius Fudge	2	651	Luna Lovegood	9	401	Slughorn	5	425
Dobby	6	613	Madam Hooch	3	52	Trelawney	5	284
Dolores Umbridge	5	663	Madame Maxime	3	201	Vernon Dursley	4	927
Draco Malfoy	16	1719	Mcgonagall	5	818	Viktor Krum	6	561
Dudley Dursley	4	477	Molly Weasley	15	83	Vincent Crabbe	7	268
Fleur Delacour	6	424	Mrs Norris	2	64			

We found evidence supporting both hypotheses. For the first hypothesis we used first-order co-occurrences which yielded an acceptable bidimensional regression coefficient. A higher-order co-occurrence like LSA yielded a high bidimensional regression coefficient, likely because its reduced sensitivity to sparsity problems of the linguistic data.

Even though narrative fictions offers a simulation of the social world around us (Mar & Oatley, 2008), the main conclusion of this study can of course be extended to the non-fictional world. We have already demonstrated this for geographical estimates for cities in the United States using newspapers (Louwerse & Zwaan, 2009), and geographical estimates for cities in the fictional Middle Earth using Lord of the Rings (Louwerse & Benesh, in press). We therefore expect that the method for the fictional Harry Potter novels can be extended to non-fictional texts. For instance, by using newspaper articles social relations among political leaders can be determined. By using blogs and tweets social networks of individuals in these texts can be estimated.

In addition to this application of the conclusion in this study, another important conclusion for the cognitive sciences is that language implicitly encodes information. In other work we have established this for geographical information (Louwerse & Benesh, in press; Louwerse, Hutchinson, & Cai, in press; Louwerse & Zwaan, 2009), bodily information (Tillman, Datla, Hutchinson, & Louwerse, in press) and other perceptual information (Louwerse, 2008; Louwerse & Connell, 2011). The current study shows that this can be extended to social information. Language has evolved such that statistical linguistic frequencies can capture the social relationships in the world around us, in the fictional world, and even in the wizarding world.

References

Backstrom, L., Boldi, P., Rosa., Ugander., & Vigna, S. (2012). Four degrees of separation. Retrieved January 31, 2012, from the arXiv database.

Berscheid, E., Snyder, M., & Omoto, A. M. (1989). The relationship closeness inventory: Assessing the closeness of interpersonal realtionships. *Journal of Personality and Social Psychology, 57,* 792-807.

Byrne, D. (1971). *The attraction paradigm*. New York, NY: Academic Press.

Ebbesen, E. B., Kjos, G. L., & Konecni, V. J. (1976). Spatial ecology: Its effects on the choice of friends and enemies. *Journal of Experimental Social Psychology*, *12*, 505–518.

Farina, A., Wheeler, D. S., & Mehta, S. (1991). The impact of an unpleasant and demeaning social interaction. *Journal of Social and Clinical Psychology*, *10*, 351–371.

Feeney, J. A., & Noller, P. (2011). Attachment style and romantic love: Relationship dissolution. *Australian Journal of Psychology*, *44*, 69–74.

Fischer, C. S. (1982). *To dwell among friends: Personal networks in town and city.* Chicago, IL: University of Chicago Press.

Friedman, A., & Kohler, B. (2003). Bidimensional regression: A method for assessing the configural similarity of cognitive maps and other two-dimensional data. *Psychological Methods, 8*, 468–491.

Hatfield, E., Traupmann, J., & Walster, G. W. (1978). Equity and extramarital sexuality. *Archives of Sexual Behavior, 7,* 127-141. Reprinted in M. Cook & G. Wilson (Eds.). (1979). *Love and attraction: An*

international conference. (pp.309-323). Oxford: Pergamon Press.

Kariam, S., Wang, D., & Leskovec, J. (2012). The life and death of online groups: Predicting group growth and longevity. *Proceedings of the ACM Conference on Web Search and Data Mining.*

Kim, J., McHale, S., Osgood, D., & Crouter, A. (2006). Longitudinal course and family correlates of sibling relationships from childhood through adolescence. *Child Development, 77*, 1746-1761.

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.) (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.

Li, W., Mollallem, I., Paller, K. A., & Gottfried, J. A. (2007). Subliminal smells can guide social preferences. *Psychological Science, 18,* 1044-1049.

Louwerse, M. M. (2008). Embodied representations are encoded in language. *Psychonomic Bulletin and Review, 15*, 838–844.

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *TopiCS in Cognitive Science, 3,* 273-302.

Louwerse, M.M. & Benesh, N. (in press). Representing spatial structure through maps and language: Lord of the Rings encodes the spatial structure of Middle Earth. *Cognitive Science*.

Louwerse, M. M., Cai, Z., Hu, X., Ventura, M., & Jeuniaux, P. (2006). Cognitively inspired natural-language based knowledge representations: Further explorations of latent semantic analysis. *International Journal of Artificial Intelligence Tools, 15,* 1021–1039

Louwerse, M. M., & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science, 35,* 381-398.

Louwerse, M. M., Hutchinson, S., & Cai, Z. (in press). The Chinese route argument: Predicting the longitude and latitude of cities in China and the Middle East using statistical linguistic frequencies. *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Louwerse, M. M. & Zwaan, R.A. (2009). Language encodes geographical information. *Cognitive Science*, *33*, 51-73.

Mar, R. A. & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, *3*, 173–192.

Muckety LLC. (2012), *Harry Potter Series* [Graphical Interactive Relationship Influence Map]. Retrieved from http://www.muckety.com/Harry-Potter-series/5017817.muckety

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM review*, *45*, 167–256.

Reis, H. T., Maniaci, M. R., Caprariello, P. A., Eastwick, P. W., & Finkel, E. J. (2011). Familiarity does indeed promote attraction in live interaction. *Journal of Personality and Social Psychology*, *101*, 557–570.

Rowling, J. K. (1998). *Harry Potter and the sorcerer's stone.* New York, NY: Scholastic Books.

Rowling, J. K. (1999a). *Harry Potter and the chamber of secrets.* New York, NY: Scholastic Books.

Rowling, J. K. (1999b). *Harry Potter and the prisoner of Azkaban.* New York, NY: Scholastic Books.

Rowling, J. K. (2000). *Harry Potter and the goblet of fire.* New York, NY: Scholastic Books.

Rowling, J. K. (2003). *Harry Potter and the order of the phoenix.* New York, NY: Scholastic Books.

Rowling, J. K. (2005). *Harry Potter and the half blood Prince.* New York, NY: Scholastic Books.

Rowling, J. K. (2007). *Harry Potter and the deathly hallows.* New York, NY: Scholastic Books.

Scott, J. (1988). Social network analysis. *Sociology*, *22*, 109–127.

Sprecher, S., & Hendrick, S.S. (2004). Self-disclosure in intimate relationships: Associations with individual and relationship characteristics over time. *Journal of Social & Clinical Psychology, 23*, 857-877

Takhteyev, Y., Gruzd, A., & Wellman, B. (2011). Geography of twitter networks. *Social Networks*.

Tillman, R., Datla, V., Hutchinson, S., & Louwerse, M. M. (in press). From head to toe: Embodiment through statistical linguistic frequencies. *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Tobler, W. R. (1964). Bidimensional regression. *Geographical Analysis, 26,* 187–212.

Touhey. J . C. (1972) Comparison of two dimensions of attitude similarity on heterosexual attraction. *Journal of Personality and Social Psychology, 23,* 8-10.

Wright, P. H., & Scanlon, M. B. (1991). Gender role orientations and friendship: Some attenuation, but gender differences abound. *Sex Roles, 24*, 551-566.

Young, F.W. (1985) Multidimensional scaling. In S. Kotz, & N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*, (Vol. 5, pp. 649-659). New York, NY: Wiley.

Zajonc, R.B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science,* 10, 224-228.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Applied Social Psychology, 9,* 1-27.