Guanxi in the Chinese Web - a Study of Mutual Linking

Valerie King
Dept. of Computer Science
University of Victoria
Victoria, BC, CANADA
val@uvic.ca

Louis Lei Yu
Dept. of Computer Science
University of Victoria
Victoria, BC, CANADA
yul@uvic.ca

Yan Zhuang [*]
Dept. of Computer Science
University of Victoria
Victoria, BC, CANADA
yzhuang@uvic.ca

ABSTRACT

Guanxi is a type of dyadic social interaction based on feelings ("*qing*") and trust ("*xin*"). Long studied by scholars of Chinese origin, it has recently drawn the attention of researchers outside of China. We define the concept of *guanxi* as applied to the interaction between web sites. We explore methods to identify *guanxi* in the Chinese web, show the unique characteristics of the Chinese web which result from it, and introduce a mechanism for simulating *guanxi* in a web graph model.

Categories and Subject Descriptors: H.4.m [Information Systems]: Information Systems Applications; J.4 [Computer Appllications]: Social and Behavioral Sciences;

General Terms: Algorithms, Theory

Keywords: web graph, social networks, stochastic graph modeling, web measurement

1. *GUANXI*

The Chinese web is notable for a large number of mutually linking web sites. We hypothesize that this is in part a manifestation of a social construct known as *guanxi*, which can be widely observed in Chinese culture. *Guanxi* has been described as "an informal ... personal connection between two individuals who are bounded by an implicit psychological contract to [maintain] a long term relationship, mutual commitment, loyalty and obligation" [2]. Dyadic relationships are the fundamental units of *guanxi* networks [2].

To establish *guanxi*, two parties must first establish a *guanxi base*: a tie between two individuals [2], e.g., same birthplace, same workplace, same family, close friendship. Also, two individuals can claim to have *guanxi* by acquaintance through a third party with whom they both have *guanxi*. Once a *guanxi* base is formed, *guanxi* can be developed through the exchange of resources ranging from moral support and friendship to favors and material goods [2].

2. *GUANXI* APPLIED TO THE WEB

We regard a web site as representing a company, a person or a news source. Two web sites may exhibit *guanxi* by mutual linking. Their linking may reflect a prior existing *guanxi* relationship, or two web sites can establish a *guanxi* base through common interests or through a third web site. We consider link exchange schemes, where only a phone call or an email is all that is required to establish the *guanxi* base and linking is done for the sole purpose of promoting one's own web site, a weaker form of *guanxi* which we call *cheap guanxi*. [1] After establishing a *guanxi* base, two web sites will reach a mutual agreement to exchange resources; in this case, these resources take the form of links. Distinguishing between strong and cheap *guanxi* is one goal of our work.

High degree nodes: As establishing strong *guanxi* takes effort, mutual links incident to nodes with many mutual links are more likely to be weak *guanxi*. In some of our studies, we filter such edges out when considering strong *guanxi*.

Triangles: If two web sites A and B establish *guanxi* via a third web site C, mutual links may form between each pairs of the web sites. We identify two structures: a *Type 1 triangle*, composed of two mutual links and one uni-directional link and a *Type 2 triangle* in which all three sides are mutual links, to be good indications of two websites establishing guanxi via a third website. Over time, we expect some Type 1 triangles to turn into Type 2 triangles. We take the number of triangles involving a mutual link to be one indication of the strength of its *guanxi*.

Textual clues: Chinese web sites often have a specially titled section of links labeled "friendly links" or sometimes in the case of commercial web sites "partnership links". These links are likely to indicate either the existence of *guanxi* or the desire to establish *guanxi* with the other web sites.

3. STRUCTURAL ANALYSIS OF *GUANXI* IN THE CHINESE WEB

We use a web graph data set which is representative of the Chinese web [4]: CWT200G collected by Peking University in May 2006 and construct a digraph as follows: each web site is represented by a node. There is a single directed edge from node A to node B in the site graph iff there is at least one link from a web page at web site A to a web page at web site B. We refer to the resulting digraph as the Chinese site graph. It has 11,570 nodes and 475,880 edges.

We randomly sampled 30,000 web sites from the data ob-

---

[1]e.g., from a travel agency web page: "Our website welcomes friendly links... .Each website's PageRank score has to be ≥ 4, ranking in Alexa $\geq 50,000$."

tained from a general web crawl conducted by Microsoft in 2006 and constructed a general site graph of 30,000 nodes and 654,240 edges. [2]

Directly comparing these two site graphs can be misleading since they are of different sizes and densities. So, we use the hostgraph model [1] (where links are created by copying links of a randomly chosen prototype node as in[3]) to generate random graphs with properties similar to the Chinese web. That is, by tuning the parameters of the hostgraph model, we randomly generate graphs comparable in size, density, and in-degree distribution to that of the Chinese site graph. We found that the hostgraph model cannot explain the unusual number of mutual links in the Chinese site graph. A detailed comparison is illustrated in Figure 1.
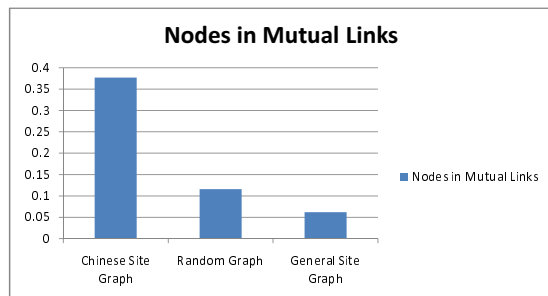

Figure 1: Fraction of nodes in mutual links in the Chinese site graph, general site graph and hostgraph model graph

4. A *GUANXI* MODEL OF THE WEB

We propose a mechanism to model the evolution of the *guanxi* structure on the web and we inject this mechanism into the hostgraph model to produce a new model for the Chinese web. The *guanxi* mechanism is defined as follows: in each time step, we add k *guanxi* edges to a node A. The destinations of the k *guanxi* edges are decided as follows: we first choose a prototype uniformly at random from the existing nodes.

1. With probability q, we add k edges with a method similar to the hostgraph model [1]. Once each edge is established, there is a probability f that the destination will link back to A.

2. With probability $1 - q$, the node A first links to the prototype and then copies the remaining $k - 1$ edges from the *guanxi* links of the prototype randomly. Once each link is established, there is a probability g that the destination will link back to A.

The copying process in (1) simulates web site A's attempt to form cheap *guanxi* links with popular web sites in order to promote his/her own web site. We set the probability f to be proportional to the relative popularity (as determined by in-degree) of A and inversely proportional to the popularity of destination B. In (2), we simulate the creation of *guanxi* links through a third party. Here g may be a fixed constant

---

[2]The crawl from Microsoft was done in a breadth-first-search fashion, after pre-processing and spam filtering, they successfully retrieved 463,685,607 HTML pages. These pages contain 17,672,011,890 hyperlinks.

if owner of both sites have established *guanxi* outside the web. Overall, the *guanxi* model can be described as follows: at each time step, depending on the density of the graph, either a new node with k edges is added or k edges are added to an existing node chosen uniformly at random. The k edges are added as follows: (1) With probability α, we add k edges to destinations using the hostgraph model; (2) With probability $1-\alpha$, we add k *guanxi* edges to destinations using the *guanxi* mechanism.

We use this new model to generate a random graph with similar properties of the Chinese site graph extracted from CWT200G. The results are summarized in Figure 2. By changing the parameters, we can control the percentage of nodes and links involved in mutual links, Type 1 and Type 2 triangles respectively.
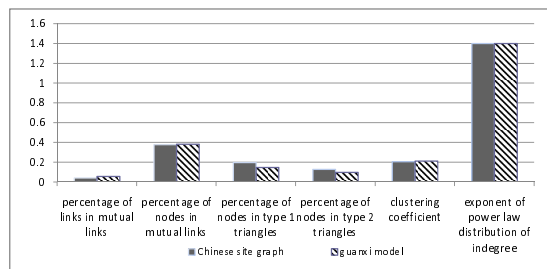

Figure 2: Simulation results

5. ONGOING WORK AND APPLICATIONS

Currently, we are conducting experiments to refine our ability to distinguish between strong and cheap *guanxi*, by analyzing textual indications of *guanxi* in the Chinese web and studying mutual links and related graph structures as they evolve over time. We are examining our findings in light of studies of social networks and the economics of link exchange schemes. To understand *guanxi* on the web as a cultural phenomon, we intend to examine site graphs of other nationalities. We believe this work may have applications to tasks such as producing personally tailored recommendations, filtering out web spam, and understanding social networks.

6. REFERENCES

[1] K. Bharat, B.-W. Chang, M. R. Henzinger, and M. Ruhl. Who links to whom: Mining linkage between web sites. In *ICDM '01: Proceedings of the 2001 IEEE International Conference on Data Mining*, pages 51–58, Washington, DC, USA, 2001. IEEE Computer Society.

[2] X.-P. Chen and C. C. Chen. On the intricacies of the chinese guanxi: A process model of guanxi development. *Asia Pacific Journal of Management*, 21(3):305–324, 09 2004.

[3] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins. The Web as a graph: Measurements, models and methods. *Lecture Notes in Computer Science*, 1627:1–17, 1999.

[4] Q. Qi. On the construction of the document set of a large scale collection – cwt200g. Master's thesis, Peking University, 2006.