Guanxi in the Chinese Web

Valerie King, Louis Yu and Yan Zhuang
Department of Computer Science
University of Victoria, Victoria, BC, Canada
Email: {*val, yul, yzhuang*}*@uvic.ca*

***Abstract*—Guanxi is a type of dyadic social interaction based on feelings, trust, and the development of friendship. In this paper, we define the concept of *guanxi* as it is applied to the interaction between web sites. We show by analyzing the local linking structure between Chinese web sites and the content of Chinese web documents that interaction between Chinese web sites can be seen to exhibit two types of *guanxi*: strong *guanxi* and cheap *guanxi*. We compare the local linking structure of Chinese web sites to the local linking structure of web sites in the general web and in Japan, Iran, and France. Finally, we explore methods to identify types of guanxi in the Chinese web, and give a method for simulating guanxi in a web graph model.**

I. INTRODUCTION

The Chinese web is notable for a large number of mutually linked web sites and other edges that we believe are associated with these links. We hypothesize that this phenomenon is a manifestation of a complex dyadic social construct in China known as *guanxi*. In this paper, we study this phenomenon between web sites in the World Wide Web, and in the Chinese portion of the World Wide Web. We give a survey of related work and introduce the concept of *guanxi*. We describe how *guanxi* is applied to the interaction between Chinese web sites and examine the local linking structure of the Chinese web to look for empirical evidence. We show that while there are linking patterns indicating the existence of underlying social ties between web site owners in the World Wide Web, these patterns are more prevalent between web sites in the Chinese web and can be explained by the logic of *guanxi*. Lastly, we classify web sites that develop types of *guanxi* based on characteristics in linking patterns and PageRank correlation. [1]

II. RELATED WORK

Our work is based on work in link analysis, web graph measurement, random graph modeling and social network analysis. We briefly survey each area.

A. Link analysis and Web Graph Measurement

A major research area in web mining is link analysis. Algorithms such as PageRank [2] are used by search engines to determine the ranking of web pages. PageRank is based on the notion of peer endorsement and is known to be highly correlated to the log of a page's in-degree [3].

Many experiments were conducted regarding the structure of web graphs [4] [5]. One important property of the general web graph is that it follows the power-law in the in-degree and out-degree distributions of nodes [6] [7] [8]. A *power-law distribution* means that the fraction of nodes with degree i is proportional to $1/i^{\alpha}$ for some constant α. Bharat et al. [9] shows that the in-degree and out-degree exponents of the general site graph are 1.62 and 1.67 respectively. Liu et al. [10] reports that the in-degree and out-degree exponents of the Chinese site graph are 1.4 and 1.5 respectively.

B. Random Graph Models

Many stochastic models have been created to generate random graphs that have attributes which resemble the web. Two main concepts have been proposed: preferential attachment [7] and copying [11]. The first evolving graph model explicitly designed to model the web was the preferential attachment model of Barabasi and Albert [7]. There, new nodes are more likely to link to existing nodes with high in-degree. Barabasi and Albert concluded that their model generates graphs whose in-degree distribution follows the power law with exponent of 3. Dorogoytsev et al. [12] generalized the Barabsi - Albert model as follows: at each time step a new site is added and m new directed links are added to both new nodes and existing nodes. Each site is assigned with an initial attractiveness score A. Over time, the attractiveness score of a site S is equal to the initial attractiveness score A plus the in-degree of S. Because this model allows for the formation of mutual links, we use this model later and refer it as the *generalized preferential attachment model*.

The first model based on the concept of copying was introduced by Kleinberg et al. [13] and was later analyzed by Kumar et al. [11]. The copying mechanism is motivated by the intuition that authors of web pages will randomly find a page and then copy some portion of the links to their own page. A notable variation of the copying model is *the hostgraph model* of Bharat et al. [9] used to model the linking structure between web sites. The hostgraph model is 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) pick a prototype uniformly at random among all existing nodes, then pick k random outgoing edges from the prototype; (2) with probability p, the destination of the i^{th} outgoing link is chosen uniformly at random; (3) with the remaining probability $1-p$, the destination of the i^{th} outgoing link is chosen to be the destination of the i^{th} outgoing link of the prototype.

[1] Some preliminary findings on this subject appeared as a poster paper in WWW2008 which was awarded best poster. [1]

Both the preferential attachment model and the copying model assume that node A links to node B because node B is of interest to node A, and this is independent of B's interest in A; we call this behavior "referencing".

C. Social Network Analysis

For many years the structural and mathematical properties of various social networks have been studied by social scientists, mathematicians and computer scientists (see [14] [15] [16] for surveys). Scott [17] identifies the various cliques, dyads, components and circles in which social networks can be formed and the significance of positions in these networks. He also surveys some theoretical and technical approaches to analyze these social networks. Mislove et al. [15] present a large-scale measurement study and analysis of the structure of online social networks such as Orkut, YouTube, and Flickr. Their results show that online social network follows the power-law and small-world properties. Software such as Pajek [18] and Ucinet 6 for Windows [19] can be used to calculate basic characteristics of social and physical networks

III. *GUANXI*

The concept of *guanxi* has been well studied in China. Even though it varies in different Chinese societies and changes over time, one can find similar notions of *guanxi* in the ancient writings of Confucius [20]. Today, *guanxi* can be observed in various areas of the popular Chinese culture. *Guanxi* has been defined as

> An informal, particularistic personal connection between two individuals who are bounded [sic] by an implicit psychological contract to follow the social norm of *guanxi* such as maintaining a long term relationship, mutual commitment, loyalty, and obligation. A quality *guanxi* is also characterized by the mutual trust and feeling developed between the two parties through numerous interactions following the self-disclosure, dynamic reciprocity, and long-term equity principles. [21]

Pablos [22] points out that the dynamic of social network is not the same in the West as it is in Confucian societies. She found that while most Western network theories focus on network structures and individual positions in the network, dyadic (two-party) relationships are the fundamental units of *guanxi* networks. Hwang [23] classified *guanxi* into three categories : (1) *Socio-affective*, which refers to family and family-like relationships; (2) *Instrumental*, which refers to the type of *guanxi* involving exchanges of resource and material needs (such as the *guanxi* between sellers and buyers of goods and services); (3) *Mix guanxi*, which both feelings and material benefits are exchanged. In the modern Chinese culture, *socio-affective guanxi* is considered to be naturally stronger than other types of *guanxi* [21].

A. The Establishment of Guanxi

To establish *guanxi*, two parties need to establish a "connection" known as a *guanxi* base [21]. One prevalent *guanxi* base consists of common social identities [24]. [2] Most frequently claimed common social identities are birthplace (*tong xiang*), educational institution (*tong xue*), workplace (tong shi), membership in the same family (*qin ren*) or close friendship (*you ren*). Two individuals can also have *guanxi* due to an acquaintance through a third party with whom they both have *guanxi* with; we call this *third party referencing*.

After a *guanxi* base has been established, to further develop *guanxi*, the two parties have to establish what are termed *qing*: feelings, which over time can turn into *xin*: trust. *Qing* and *xin* can be developed through the exchange of resources from both parties. The resources exchanged can range from moral support and friendship to favors and even material goods. The exchange of resources is said to follow *the general principle of reciprocity* [21]: first, there is no specification as to when a person should return a received favor, rather, the exchange happens over time; second, the amount of exchange can be somewhat flexible. Chung and Hamilton [25] states "there is no way to quantify the favor in order to find out whether they are of equal value ... however, both parties need to perceive them more or less as equal."

B. Dyadic Relationships in General Social Networks

Although *guanxi* refers to dyadic relationships with specific properties that occur in the Chinese society, dyadic relationships in general occur in many cultures [26]. Wasserman et al. [27] documented the general structure and mathematical properties of dyads in social networks.

Social capital is a concept in sociology referring to connections within and between social networks. In *The Forms of Capital* [28], Bourdieu distinguishes between four different types of capitals. Among them *economic capital* refers to cash and assets, and *social capital* refers to resources based on group membership, relationships, networks of influence and support. We see *guanxi* as the exchange of *social capital*. Our view is supported by the work of Lin et al. [29], who talks about the exchange of what they call *guanxi capital*.

In mathematical sociology, *interpersonal tie/social tie* [30] [31] is a general term describing the connection and relationship between people. Social ties come in three varieties: strong, weak, or absent. It is documented that in society, the majority of the ties in social networks are weak ties [32]. We consider *guanxi* a particular kind of social tie.

IV. *GUANXI* APPLIED TO THE WEB

Regarding a web site as representing a company, a person or a news source, we propose that in the Chinese web, two web sites may exhibit *guanxi*, i.e., their linking may reflect an existing *guanxi* relationship.We call this *web based guanxi*.

Before establishing *guanxi*, two web sites must first establish a *guanxi* base. Some common *guanxi* bases include: (1) web sites that belong to members of the same family or close friends; (2) web sites that belong to collaborating partners; (3) web sites that originate in the same province or city; (4) web

[2] In the Chinese language, the word "*tong*" means same, common or shared.

sites that have the same purpose or subject; (5) web sites that belong to a friend of a friend.

After establishing a *guanxi* base, two web sites may establish *guanxi* and thereby reach a mutual agreement to exchange resources. In this case, these resources take the form of links, resulting in the establishment of mutual links over time. As a web site may send a link through one page and receive a link back to another page, we consider mutual links between web sites rather than merely mutual links between web pages.

Some *guanxi* between web sites are naturally stronger than others. *Guanxi* bases (1) and (2) from the above list can form naturally stronger *guanxi*, while (3) - (5) can form naturally weaker *guanxi*. In some cases, a phone call or an email is all that is required to establish a *guanxi* base which proceeds with the exchange of links. The relationship requires little effort and feeling and is done for the sole purpose of establishing links to promote one's own web site, We consider this superficial or cheap *guanxi*. It is not uncommon to see Chinese web sites inviting "friendly links" on their home pages with the idea that these links will be reciprocated. [3]

We speculate that Chinese site owners wish to establish cheap *guanxi* for two reasons. First, if site A establishes cheap *guanxi* with site B, A's owner believes that users visiting B may visit A by clicking on B's link to A. Second, if we are given web site i and web site j with PageRank $P(i)$ and $P(j)$ at time t, then by using the perturbation bound (Theorem 2.1) provided by Cho and Meyer [33] we can calculate the change in i's PageRank after a mutual link had been added between i and j at time $t+1$. The perturbation bound expresses the relative change in i's PageRank by the maximum *mean passage time* between node i and any node $n \neq i$ in the graph. After a mutual link had been established between i and j, the maximum *mean passage time* can change due to the addition. Thus node i and node j may lose or gain PageRank if they establish a mutual link. We expect that for web sites establishing cheap *guanxi*, the objective is to gain PageRank through the establishment. Furthermore, Theorem 1 of [34] implies that $P(j)-P(i)$ is correlated to the amount of PageRank node i will gain or lose if i establishes a mutual link with node j. We believe that Chinese web sites generally prefer to establish cheap *guanxi* with web sites of higher PageRank. However, since neither web site is willing to lose PageRank from the establishment of a mutual link, we speculate mutually linked web sites tend to have similar PageRank scores [4].

The need for Chinese web sites to establish cheap *guanxi* gave rise to link exchange platform web sites such as *www.link-8.com*, and *www.qqfc.com.cn*.[5] We see these link exchange platform web sites as a type of *guanxi* base.

We hypothesize some local linking patterns that may result from web sites developing cheap *guanxi* with each other. Suppose web site A and web site B developed cheap *guanxi*. Another web site C can develop cheap *guanxi* with both A and B through a link exchange platform website. Alternatively, C that has already developed cheap *guanxi* with A may look at A's outgoing links looking for potential partners to develop cheap *guanxi* with (third party referencing). We speculate that a large amount of *Type 1 triangles*, composed of two mutual links and one uni-directional link, and *Type 2 triangles*, in which all three sides are mutual links, will result from web sites developing cheap *guanxi* with each another. Over time, a *Type 1 triangle* developed around a *guanxi* link can develop into a *Type 2 triangle*. [6]

A. Textual Indication of Guanxi

One prominent clue to the existence of *guanxi* links is textual. Chinese web sites often have a specially titled section of links labeled "friendly links"(*you qing lian jie*) or sometimes in the case of commercial web sites "partnership links" (*he zuo huo ban*). These labeled sections are likely to indicate either the existence of *guanxi* or the desire to establish *guanxi* with other web sites. The literary translation of *you qing lian jie* is linking to friends. The literary translation of *he zuo huo ban* is collaborating partners.

B. Mutual Links Between General Web Sites

Dyadic relationships between web sites also occur among non-Chinese web sites. Mutual links between friends seem common between blogs [35]. A link exchange scheme [36] is described as the process of exchanging text links or banner links between web sites in order to promote each others' web sites. Gyongyi and Garcia-Molina [37] and Du et al. [38] identified the structure of an optimal spam farm. We believe *guanxi* between Chinese web sites is much more than mere link exchange schemes; it is a natural extension of a type of social network widely accepted in the Chinese society. It comes naturally to most Chinese web masters to promote web sites which belong to people they have *guanxi* with. It appears in a large variety of web sites, personal or commercial and our evidence suggests that it is a much more prominent feature of the Chinese web structure as compared to the general web structure. We observe that Chinese web sites do not follow the optimal spamming structure [37] [38] when developing *guanxi* with one another.

[3] e.g., from a travel agency web page: " Our website have PageRank score of 4 ... we welcomes friendly links. There are no requirements on content or purposes to link to our page; however, the web site has to be legal, with no viruses ... Each web site's PageRank score has to be ≥ 4."

[4] As a textual indication, we often see Chinese web sites specifying on their web pages that they will only exchange links with web sites of the same PageRank or higher.

[5] The instruction on www.link-8.com states "... to establish *guanxi*, follow the three simple steps. First, initiate a link to the web site you wish to establish *guanxi* with. Second, contact the site owner through telephone or instant messenger. Third, the link should be recirpocated in a few days."

[6] We initially hypothesized in our work [1] that *Type 1 triangles* and *Type 2 triangles* are the result of web sites developing strong *guanxi* with each other. However, after some initial experimentation, we observed that the triangles might be a result of web sites developing cheap *guanxi* with each other.

V. EMPIRICAL STUDY OF THE CHINESE WEB AND THE GENERAL WEB

In this section, we examine the local linking structure of the Chinese web and the general web to determine if there is structural evidence of *guanxi*.

A. Mutual Links, Type 1 and Type 2 Triangles

We randomly selected 10,000 web sites from a crawl of the Chinese web conducted by Peking University's Sky Net search engine in 2007. We refer to these 10,000 web sites as the sample Chinese web. We also randomly selected 20,000 web sites from the Open Directory Project (*www.dmoz.com*) as the sample general web. We use the uniform random sampling method to select web sites that represent the Chinese web and the general web. Becchetti et al. [39] shows that the fraction of reciprocated links in a small set of nodes (10%) chosen from a large graph using the uniform random sampling method is similar to the fraction of reciprocated links in the large graph.

We crawled the web sites in both samples and their neighboring web sites reachable by outgoing links. Although the sub-graph obtained this way does not match the in-degree distribution of the original graph [40], it is sufficient for us to find all the mutual links and *Type 2 triangles* involving web sites in both samples. We examined the percentages of sites involved in mutual links and *Type 2 triangles*, and estimated the percentage of sites involved in *Type 1 triangles*, in both the sample Chinese web and the sample general web. We also calculated the total number of mutual links as a percentage of the total number of outgoing links in both samples.

We found that the percentage of web sites and their corresponding outgoing links involved in mutual links, *Type 1 triangles* and *Type 2 triangles* in the sample Chinese web are significantly higher than those in the sample general web. However, this can be expected for two reasons. First, the Chinese web is relatively denser than the general web, i.e., with a higher ratio of edges to nodes. According to the experiment conducted by Bharat et al. [9], the indegree and outdegree exponents of the general site graph are 1.62 and 1.67. The in-degree and out-degree exponents of the Chinese site graph as reported by Liu et al. [10] are 1.4 and 1.5 respectively. Second, the Chinese web has been historically smaller. Even if we presume that links are created uniformly at random, a graph with a smaller node set and the same density is likely to have a higher proportion of mutual links and triangles. So if we presume that the Chinese web is quite isolated from the rest of the general web, a higher fraction of links in mutual links and triangles are expected from its size and density alone.

In order to determine if there is a high percentage of nodes and links involved in mutual links, *Type 1 triangles* and *Type 2 triangles* in the Chinese web, we first need to understand what to expect from a web of similar size and density. To do so, we use a graph model to generate a random graph with size and density similar to the Chinese web and compare our data to the predictions given by this model. We know of no stochastic models that generate random graphs which represent every aspect of the World Wide Web. We chose the hostgraph model [9] to generate our random graph because the hostgraph model (1) is used specifically to model the link structure between web sites; (2) allows the emergence of mutual links; (3) generates random graphs with bipartite cliques of "hubs" and "authorities" which have been found to be characteristic of the web [5]. Using the method by Zhuang [41], we implemented the hostgraph model [9] and generated a random graph comparable in size and density to that of the Chinese web in 2007. We examined the percentage of nodes that are involved in mutual links, *Type 1 triangles* and *Type 2 triangles* and the percentage of links that are involved in mutual links in the random graph.

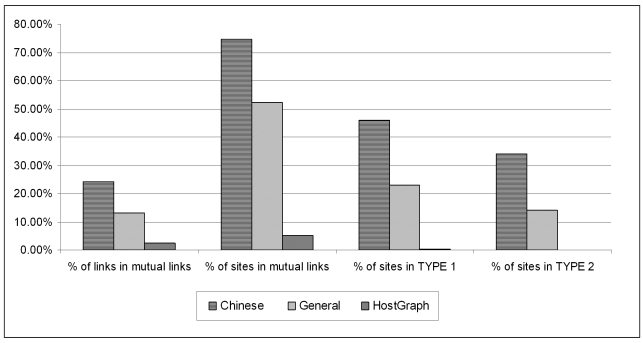


Fig. 1. Mutual Links and Triangles in the Sample Chinese and General Web

Figure 1 shows the experimental results from the sample Chinese web, the sample general web and the random graph generated by the hostgraph model. The overall results indicate that mutual linking widely exists in the World Wide Web, and is especially significant in the Chinese web, and it is not predicted by the hostgraph model.

Garlaschelli and Loffredo [42] looked at the non-random presence of mutual links between pairs of vertices. A new measure of *link reciprocity* was proposed which can be used to compare the number of mutual links whose formation is not attributed to chance. Using their measure of *link reciprocity* p [7], we found $p = 0.226$ for the sample Chinese web and $p = 0.127$ for the sample general web. The sample Chinese web has a much higher degree of *link reciprocity*.

B. PageRank Correlation

In order to demonstrate that mutual linking is a possible indication of *guanxi* in the web, and not the result of referencing under the *generalized preferential attachment model* [12] or the *hostgraph model* [9], we examine correlations between the sites' PageRank scores. We look at two types of correlations: PageRank correlation between web sites that are connected by single links C_S, and PageRank correlation between web sites that are connected by mutual links C_M. We define C_S and C_M as follows:

[7] $p = \frac{1-\overline{a}}{r-\overline{a}}$, where r is the number of mutual links as a percentage of the total number of links in the graph, and $\overline{a}$ is the ratio of observed to possible directed links (link density).

Let E be a set of pairs (x, y) such that $(x, y) \in E$ iff site x points to site y

$$S_i = \{(x, y) \mid (x, y) \in E \text{ and } (y, x) \notin E \text{ and } PageRank(x) = i\}$$

$$M_i = \{(x, y) \mid (x, y) \in E \text{ and } (y, x) \in E \text{ and } PageRank(x) = i\}$$

$$C_S(i) = \frac{\sum_{(x,y)\in S_i} PageRank(y)}{|S_i|}$$

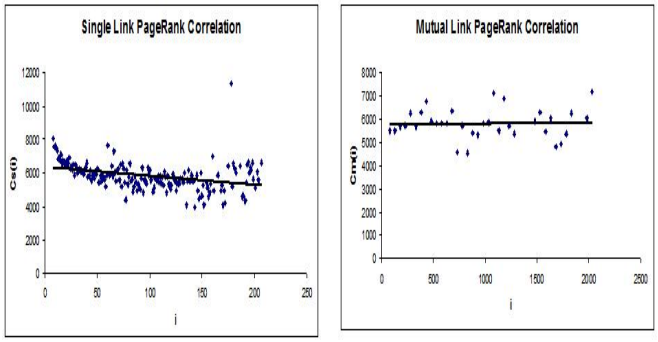
$$C_M(i) = \frac{\sum_{(x,y)\in M_i} PageRank(y)}{|M_i|}$$

$C_S(i)$ ($C_M(i)$) is the average PageRank of nodes with single (mutual) links from nodes with PageRank equal to i. Under both the *generalized preferential attachment* model [12] and the *hostgraph model* [9], the destinations of all links are determined independently of the PageRank of the source nodes. The PageRank correlation between singly connected and mutually connected nodes in such graphs is nontrivial, and we ran simulations to determine this. We generated a random graphs of 50,000 nodes and 500,000 links using the *generalized preferential attachment* model [12]. We calculated the PageRank scores for nodes in the graph using the PageRank algorithm of Brin and Page [2]. Figure 2(a) and 2(b) show the PageRank correlation between nodes connected by single links and by mutual links in the random graph. We see that $C_S(i)$ fluctuates mostly in the range of 4000×10^{-6} to 8000×10^{-6}, and $C_M(i)$ fluctuates mostly in the range of 4000×10^{-6} to 7000×10^{-6}; the plots of $C_S(i)$ and $C_M(i)$ appear to be flat. This indicates that the PageRank i of the source nodes has little impact on $C_S(i)$ or $C_M(i)$.

We then looked at PageRank assigned to web sites in the sample Chinese web and the sample general web by the Google search engine [8]. We calculated PageRank correlations between singly connected and mutually connected nodes (Figures $3(a)$ and $3(b)$). In the case of single links, the plot of $C_S(i)$ as a function of PageRank i is almost flat, which agrees with the prediction of the *generalized preferential attachment* model [12] and the *hostgraph model* [9]. But in both the sample Chinese web and the sample general web, the PageRank values between mutually linked web sites (the plot of $C_M(i)$) exhibit a very strong positive correlation, i.e. sites with high PageRank are more likely to be mutually linked to sites with high PageRank, and sites with low PageRank are more likely to be mutually linked to sites with low PageRank. A similar phenomena has been called assortative mixing [43], and has been discovered in many undirected collaboration networks such as movie co-starring network [43], and paper co-authorship network [43]. This positive correlation gives strong evidence that mutual linking in the web represents a type of interaction different from referencing.

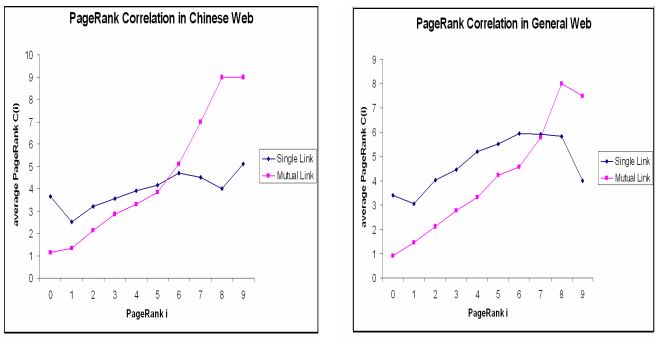
In the theory of *guanxi*, such positive correlation of PageRank can be explained by the logic of reciprocity [25] (see Section III-A) where both parties are expected to exchange similar amount of resources. In the case of web sites, the resources are mainly user traffic and PageRank value (see Section IV). In Figure $4(a)$, we measure the absolute value of the PageRank difference distribution between mutually linked web sites in the sample Chinese web. The result supports our hypothesis that mutually linked web sites follow the logic of reciprocity. We note that the fraction of two web sites mutually linked decreases exponentially as the absolute value of PageRank difference increases.

In order to further discover how mutual links are established between web sites, we measure the average number of mutual links of sites with a particular PageRank in the sample Chinese web (Figure $4(b)$). We observe that the average number of mutual links decreases steadily after it reaches its maximum value when PageRank equals 4. (Note that we do not consider the part with PageRank > 8 since that we have too few samples in that range to give an accurate estimate). One possible explanation of this phenomenon is that web sites with more resources (high PageRank) are more attractive to other web sites that wish to establish cheap *guanxi*. But, due to the logic of reciprocity [25] (see Section III-A), web sites



(a) Single Link (b) Mutual Link

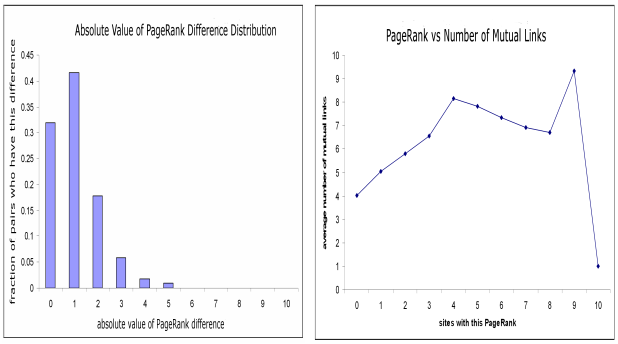
Fig. 2. PR Correlation in the GPA (scaled by a factor of 10^6)



(a) (b)

Fig. 3. PageRank Correlation in the Sample General and Chinese Web

[8]PageRank is provided to users via the Google toolbar and Google directory. When a page is visited, the toolbar lists its PageRank on a scale of 0 to 10



(a) (b)

Fig. 4. PR Difference Distribution and PR vs Number of Mutual Links

generally prefer to respond with a link back to those web sites that have at least similar amount of resources (see Section IV). If we describe this mathematically, let p be the PageRank of a web site and let $f(p)$ be the attractiveness of the web site where $f(p)$ increases as p increases. Let $r(p)$ be the probability that the web site will respond where $r(p)$ decreases as p increases. We can therefore write the ability of the web site with PageRank p to obtain new mutual links, $A(p)$, as the product of functions f and r: $A(p) = f(p) * r(p)$. Our observation implies that between PageRank 0 and 8, there exists a threshold t which makes $f(p)$ dominant when $p < t$ and $r(p)$ dominant when $p > t$.

C. A Guanxi Model of the Web

We have demonstrated that the usual random graph models do not explain *guanxi*. We propose a mechanism to model the formation of *guanxi* structure in the web. We inject this mechanism into the hostgraph model to produce a new model for the Chinese web [9]. 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 existing nodes; (1) with probability q, we establish k edges with a method similar to the hostgraph model [9]. 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 copy 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* with other web sites. In this case, we hypothesize that the probability source site A receives a link back from destination site B is proportional to A's PageRank [2] P_A and is inversely proportional to B's PageRank P_B. We set $f = \frac{d_1 P_A}{P_A + P_B}$, where d_1 is a control factor. In (2), we simulate the establishment of *guanxi* links through a third party (the formation of triangles) where the source establishes *guanxi* links with nodes that the prototype has *guanxi* with after the establishment of a *guanxi* link between the source and the prototype. We set the probability g in (2) to $g = \frac{d_2 P_A}{P_A + P_B}$ where d_2 is another control factor.

[9]See a preliminary version of this mechanism in [1]

Parameter	Chinese	General
α	0.37	0.61
q	0.44	0.71
d_1	0.87	0.82
d_2	0.83	0.74

TABLE I
SETTINGS OF SIMULATION

We inject the *guanxi* mechanism into the hostgraph model [9] 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: with some fixed probability α, we add k edges to destinations using the hostgraph model. With remaining probability $1 - \alpha$, we add k *guanxi* edges to destinations using the *guanxi* mechanism.

We use this new model to generate random graphs with similar properties as the sample Chinese web and the sample general web. Table I shows the parameters used to generate each random graph. Parameter α is set lower to generate the random graph that represents the sample Chinese web. This means that the probability for a node to establish *guanxi* $(1 - \alpha)$ has to be set higher to generate a graph that has similar properties as the sample Chinese web. Similarily the parameter q is set lower (which means $(1-q)$, the probability for a node to form triangles, is set higher) to generate the random graph that represents the Chinese web.

D. Comparison Against Web Sites in Different Countries

As a natural extension of the empirical study in Section V-A, we look at dyadic relationships between web sites in different countries. We conducted the following preliminary experiment: we chose three countries: Japan, France and Iran. For each country, we randomly selected 5000 web sites from the Open Directory Project (*www.dmoz.com*). We use the following method to determine whether a web site is from these three countries: (1) if the content of the web pages is written in French, Japanese, or Persian; (2) if the web site's url is of .fr (France), .jp (Japan), or .ir (Iran) domain.

We compared the percentage of web sites involved in mutual links, *Type 1 triangles*, and *Type 2 triangles* and the percentage of links involved in mutual links for these three countries and the sample Chinese web. Figure 5 illustrates this comparison. We see that the percentage of Chinese web sites involved in mutual links, *Type 1 triangles*, and *Type 2 triangles* and the percentage of corresponding outgoing links involved in mutual links are the highest compared to web sites from other countries and their corresponding outgoing links. More experiments are needed to explain the presence of mutual links, *Type 1 triangles*, and *Type 2 triangles* involving web

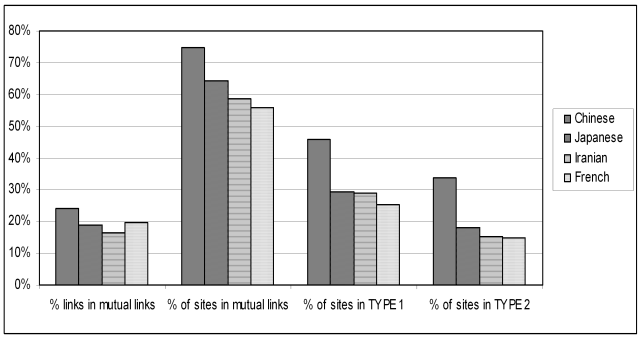


Fig. 5. Sites and Links in Different Countries in Mutual Links and Triangles

sites in other countries.

VI. INDENTIFYING *GUANXI* WEB SITES

In this section, we identify some characteristics in local linking patterns and PageRank correlation that can be used to separate web sites that establish cheap *guanxi*, strong *guanxi* and no *guanxi*. We randomly selected 17,461 web sites from the Open Directory Project (*www.dmoz.com*). For each web site, we identify web pages containing keywords indicating the establishment of *guanxi* such as *you qing lian jie* or *he zhuo huo ban*. We also search for keywords indicating the establishment of cheap *guanxi* such as "link exchange" (*jiao huan lian jie*) . We separate the web sites into three groups: 2372 web sites containing textual clues indicating they are establishing strong *guanxi*, 4938 web sites containing textual clues indicating they are establishing cheap *guanxi*, and 10151 web sites that do not contain these textual clues. We call these three groups of web sites *textual strong guanxi* web sites, *textual cheap guanxi* web sites and *textual no guanxi* web sites, respectively.

A. Mutual Links of Textual Web Sites

We found that almost all of the *textual strong guanxi* web sites and *textual cheap guanxi* web sites establish mutual links with neighbouring web sites. Only 12% of the *textual no guanxi* web sites are involved in mutual links.

We also found that on average 45% of the *textual cheap guanxi* web sites' outgoing links are reciprocated. 25% of the *textual strong guanxi* web sites' outgoing links are reciprocated. Only 8% of the *textual no guanxi* web sites' outgoing links are reciprocated. We manually inspected 100 *textual cheap guanxi* web sites and found that most nonreciprocal outgoing links are to large web sites such as *Google*, *Baidu*, or *Yahoo*, which in turn do not link back.

B. PageRank Difference in Textual Web Sites

We looked at the difference in PageRank scores between textual web sites and their neighbouring web sites connected by mutual links. We found that for *textual cheap guanxi* web sites, on average 67% of the neighbouring web sites connected by mutual links have the PageRank difference of one or zero. For 98% of the *textual cheap guanxi* web sites, greater than 50% of the neighbouring web sites connected by mutual links have the PageRank difference of one or zero.

For *textual strong guanxi* web sites, on average 36% of the neighbouring web sites connected by mutual links have the PageRank difference of one or zero. For 90% of the *textual strong guanxi* web sites, $30\% - 50\%$ of the neighbouring web sites connected by mutual links have the PageRank difference of one or zero.

For *textual no guanxi* web sites, on average 15% of the neighbouring web sites connected by mutual links have the PageRank difference of one or zero. For 87% of the *textual no guanxi* web sites, less than 30% of the web sites connected by mutual links have the PageRank difference of one or zero.

C. Identifying Types of Textual Web Sites in the Chinese Web

We hypothesize that we can use the observations mentioned in Sections VI-A and VI-B to identify web sites that are developing strong *guanxi*, cheap *guanxi* and no *guanxi*. After some experimentation with different parameters, we formulated some specific rules that can be used to separate these web sites. To identify cheap *guanxi* web sites, we look for web sites with $> 30\%$ of their outgoing links reciprocated, or web sites that have $> 50\%$ of the neighbouring web sites connected by mutual links with PageRank difference of one or zero. To identify strong *guanxi* web sites, we look for web sites with $10\% - 18\%$ of their outgoing links reciprocated, or web sites that have $30\% - 50\%$ of the neighbouring web sites connected by mutual links with PageRank difference of one or zero. To identify no *guanxi* web sites, we look for web sites with $< 10\%$ of their outgoing links reciprocated, or web sites that have $< 30\%$ of the neighbouring web sites connected by mutual links with PageRank difference of one or zero.

Using the above rules, we identify three groups of web sites, we call them *structural cheap guanxi*, *structural strong guanxi* and *structural no guanxi* web sites. We look at the correlation between *textural guanxi* web sites and their corresponding *structural guanxi* web sites. We observe that more than 77% of the *textural cheap guanxi* web sites are *structural cheap guanxi* web sites, more than 79% of the *textural strong guanxi* web sites are *structural strong guanxi* web sites and more than 90% of the *textural no guanxi* web sites are *structural no guanxi* web sites. This indicates that the *textural guanxi* web sites and their corresponding *structural guanxi* web sites are closely related. However, the correlations are still lower than we expected. We hypothesize that this is due to errors in selecting keywords. We manually inspect 100 *structural cheap guanxi* web sites, 100 *structural strong guanxi* web sites and 100 *structural no guanxi* web sites. We found 89% of the *structural cheap guanxi* web sites are establishing cheap *guanxi*, 94% of the structural strong guanxi web sites are establishing strong *guanxi* and 91% of the structural no guanxi web sites are not establishing *guanxi*.

VII. CONCLUSIONS

We defined *guanxi* in the web: particular link patterns that appear in the web as well as supporting textual evidence in

web pages which we believe are indicative of the presence and varying strengths of the underlying *guanxi* between web site owners. Through empirical study of the Chinese web, the general web, and the Japanese, Iranian, and French web, we found link patterns which indicate *guanxi* are more prevalent between web sites in the Chinese web. In particular, the Chinese web has a higher percentage of web sites engaged in mutual linking and a higher percentage of links which are part of mutual links. Ande, more Chinese web sites are involved in the triangle structures associated with cheap *guanxi*.

We studied PageRank correlation between web sites connected by single links and web sites connected by mutual links, in both the Chinese web and the general web. We found a very strong positive correlation between mutually linked web sites. We showed that such a correlation is not found between web sites connected by single links, and this positive correlation is not explained in the preferential attachment model. Not only do web site owners appear to be engaging in deliberate mutual linking, but the positive correlation of PageRank suggests that this behavior follows the resource exchange rule of *guanxi*. Based on our findings, we present a mechanism to model the *guanxi* structure in the web.

Our work is a study of how the web reflects the underlying structure of society. While it is perhaps impossible to prove the existence of *guanxi* in the web, we believe that we have made a persuasive case for the existence of a special mutual relationship between web sites which is indicated by linking patterns and textual clues. Our hypothesis is supported by our finding that these relationships appear more frequently in the Chinese web. In a culture in which the dyadic relationship of *guanxi* is so important, we believe that this phenomenon is interesting in itself, and it may be useful for producing personally tailored recommendations, filtering out web spam, and understanding social networks.

REFERENCES

[1] V. King, L. L. Yu, and Y. Zhuang, "Guanxi in the chinese web - a study of mutual linking," in *WWW '08*, pp. 1161–1162.

[2] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," in *WWW '98*, pp. 107–117.

[3] T. Upstill, "Predicting fame and fortune: Pagerank or indegree," in *ADCS'03*, pp. 31–40.

[4] R. Albert, H. Jeong, and A.-L. Barabasi, "The diameter of the world wide web," *Nature*, vol. 401, p. 130, 1999.

[5] J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "The web as a graph: Measurements, models, and methods," *Computing and Combinatorics*, pp. 1–17, 1999.

[6] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "Trawling the web for emerging cyber-communities," in *WWW '99*, pp. 1481–1493.

[7] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, p. 509, 1999.

[8] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph structure in the web," in *WWW '00*, pp. 309–320.

[9] K. Bharat, B.-W. Chang, M. R. Henzinger, and M. Ruhl, "Who links to whom: mining linkage between web sites," in *ICDM '01*, pp. 51–58.

[10] G. Liu, Y. Yu, J. Han, and G.-R. Xue, "China web graph measurements and evolution," in *APWeb'05*, pp. 668–679.

[11] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, "Stochastic models for the web graph," in *FOCS '00*, p. 57.

[12] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, "Structure of growing networks with preferential linking," *Phys. Rev. Lett.*, vol. 85, no. 21, pp. 4633–4636, Nov 2000.

[13] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[14] M. Jamali and H. Abolhassani, "Different aspects of social network analysis," in *WI'06*, pp. 66–72.

[15] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *IMC '07*, pp. 29–42.

[16] M. Buchanan, *Nexus: Small Worlds and the Groundbreaking Theory of Networks*. W. W. Norton & Company, May 2003.

[17] J. P. Scott, *Social Network Analysis: A Handbook*. SAGE Publications, January 2000.

[18] W. Denooy, A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis with Pajek*. Cambridge University Press, January 2005.

[19] S. P. Borgatti, M. G. Everett, and L. C. Freeman, "Ucinet 6 for windows: software for social network analysis," Analytic Technologies, 2002.

[20] L. Jacobs, G. Guopei, and P. Herbig, "Confucian roots in china: a force for today's business," *Management Decision*, vol. 33, no. 10, pp. 29–34, 1995.

[21] 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*, vol. 21, no. 3, pp. 305–324, September 2004.

[22] P. O. Pablos, "Western and eastern views on social networks," *The Learning Organization: An International Journal*, vol. 12, no. 5, pp. 436–456, May 2005.

[23] K. Hwang, "Face and favor: The chinese power game," *American Journal of Sociology*, vol. 92, no. 4, pp. 944–974, 1987.

[24] E. A. Langenberg, *Guanxi and Business Strategy*, ser. Contributions to Management Science. Physica, 2007, pp. 27–118.

[25] W.-K. Chung and G. G. Hamilton, *Rules and Networks: the Legal Culture of Global Business Transactions*. Oxford, Hart Publishing, 2001, ch. 11, pp. 325–346.

[26] M. Mauss, *The Gift: The Form and Reason for Exchange in Archaic Societies*. W. W. Norton & Company, August 2000.

[27] S. Wasserman, K. Faust, and D. Iacobucci, *Social Network Analysis : Methods and Applications*. Cambridge University Press, November 1994.

[28] P. Bourdieu, *Handbook of theory and research for the sociology of education*. New York: Greenwood, 1985, ch. 9, pp. 241–258.

[29] N. Lin, R. S. Burt, and K. Cook, *Social Capital: Theory and Research*. Aldine de Gruyter,US, 2001, ch. 11, pp. 278–293.

[30] A.-L. Barabasi, *Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life*. Plume Books, April 2003.

[31] M. Granovetter, "The impact of social structure on economic outcomes," *Journal of Economic Perspectives*, vol. 19, no. 1, pp. 33–50, Winter 2005.

[32] ——, "The strength of weak ties," *AJS*, vol. 78, no. 6, pp. 1360–1380, 1973.

[33] G. E. Cho, G. E. Cho, C. D. Meyer, and C. D. Meyer, "Markov chain sensitivity measured by mean first passage times," *Linear Algebra Appl*, vol. 316, pp. 21–28, 2000.

[34] K. Avrachenkov and N. Litvak, "The effect of new links on google pagerank," *Stoch. Models*, vol. 22, p. 2006, 2006.

[35] R. Kumar, J. Novak, P. Raghavan, and A. Tomkins, "Structure and evolution of blogspace," *Commun. ACM*, vol. 47, no. 12, pp. 35–39, 2004.

[36] B. Wu and K. Chellapilla, "Extracting link spam using biased random walks from spam seed sets," in *AIRWeb '07*, pp. 37–44.

[37] Z. Gyöngyi and H. Garcia-Molina, "Link spam alliances," in *VLDB '05*, pp. 517–528.

[38] Y. Du, Y. Shi, and X. Zhao, "Using spam farm to boost pagerank," in *AIRWeb '07*, pp. 29–36.

[39] L. Becchetti, C. Castillo, D. Donato, and A. Fazzone, "A comparison of sampling techniques for web characterization," in *LinkKDD'06*, August 2006.

[40] J. Leskovec and C. Faloutsos, "Sampling from large graphs," in *KDD '06*, pp. 631–636.

[41] Y. Zhuang, "Efficient sampling from random web graphs and its application," Master's thesis, University of Victoria, 2008.

[42] D. Garlaschelli and M. I. Loffredo, "Patterns of link reciprocity in directed networks," *Physical Review Letters*, vol. 93, no. 26, p. 268701, 2004.

[43] M. E. J. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, p. 208701, 2002.