Calculating Popularity Using a Simple Algorithm

Gananath R and Sreenath R

Corresponding author Email: gananathr@gmail.com

Abstract

The quality of being liked or accepted by peoples is called as popularity. Popularity is calculated by different methods according to the area of application. To calculate the popularity of websites in internet; methods like 'Click Popularity' and 'Link Popularity' were used. Developmental Psychologist uses sociometric tests to calculate popularity among peer groups. In this paper we have formulated a simple algorithm through which popularity of a candidate can be calculated and also we carried out monte carlo simulation studies of the algorithm. Popularity of either living or nonliving (eg: a commercial product) candidates can be measured using this algorithm.

Keywords: Popularity, Monte carlo, Ranking, Algorithm

Mathematics Subject Classification (2010): 00- 00A99

Introduction

Popularity is the state of being liked, enjoyed or accepted by a large number of people [7]. 'Popularity' word is originated from Latin word *popularis* which meant common or well liked. Popularity generally involves respect in two way directions; a popular person is respected by peers and will reciprocate that respect. Opinion polls are one way to rate popularity.

Internet is used by millions of people for their social, informational and consumer needs; and most of the user's use websites which are already popular in the web. There are two main ways to measure the website popularity in World Wide Web; 'Click Popularity' and 'Link Popularity'. Click popularity is the frequency with which users have visited or clicked on a site. Some search engines like direct hit uses this technique to rank website according to popularity but it is subject to artificial marketing manipulations. Link popularity or peer reviewed popularity which is less likely to be manipulated relies on links from sites to other sites rather than on usage. The websites identified using link popularity is found to have high quality websites comparing to click popularity [3]. Google which is one of the most popular search engine use page rank algorithm which works on the principle of link popularity for ranking results.

$$R(u) = c \sum_{v \in B_u} \frac{R(v)}{N_v}$$

Where R*(u)* is the popularity score associated to page u, N_v is the number of web links found in the page v, c is a constant and B_u is the set of pages to page u [5,6]. Internet is now a commonly used tool for our daily activities and web users uses internet to increase their own

popularity. Studies found out that introverts and extroverts use social networking sites such as facebook to increase their own popularity [4].

Popularity of the peer group in childhood and adolescence were studied by developmental psychologists who are interested in the social structure and dynamics. Sociometric popularity and perceived popularity are two main type of popularity assessed by developmental psychologists. Traditionally, the study of peer relations has focused on sociometric status, referred to peoples who had a lot of friends, were well-liked by many and disliked by few, who were nice to others, and who were on an even keel emotionally. Sociometric popularity is usually assessed with a peer nomination procedure, in which participants are asked to name the peers in their grade or group who they like most and like least [1]. Perceived popularity, on the other hand was assessed by asking adolescents to nominate up to three classmates they considered popular [2].

We formulated a simple algorithm which can be used to measure the popularity of a candidate. Calculating popularity of both animate and inanimate candidates can be carried out using this algorithm.

Algorithm

A candidate is to whom for which popularity should be calculated and others who are related to the candidate (directly or indirectly) are called as individuals and there relations to the candidate is called as paths (figure: 1) .To calculate the popularity of a candidate the path scores (P) for the candidate should be known and it can be calculated by initially finding out $Path^{max}$. $Path^{max}$ is the maximum number of individuals in a path of the candidate. Paths score can be calculated by taking the ratio of number of individuals in a path to $path^{max}$.

$$Path\ Score\ (P) = \frac{No: of\ Individuals\ in\ a\ Path}{Path^{max}}$$

The total score (T) can be calculated by the addition of the path scores.

$$T=P_1 + P_2 + P_3 + \cdots + P_n$$

$$T = \sum_{i=1}^{n} P_i$$

Where *n* is the number of paths.

The Log Score (L) of the candidate can be calculated by taking logarithm of the total score T.

$$L = log_{10}\, T$$

If L is high then the candidate is highly popular and if it's low then the candidate is less popular.

An example

From the figure 1; to calculate the popularity of candidate (CA) we have to first measure its path scores. There are four paths for CA they are Path 1, 2, 3 and 4.

for example in path 1.

We assume that candidate (CA) is the son of *d*, *a* be the friend of *d*, *b* be the spouse of *a* and *c* be the co-worker of *b*; *a*, *b* and *c* knew CA through *d*, *a* and *b* respectively. Path 1 has the highest individuals 4 (a, b, c & d), so path 1 is the Path^{max}. Below are the four paths and number of individuals in each path.

$$Path_1^{max} = 4\ (a, b, c\ , d);\ Path_2 = 3(a, b, c);\ Path_3 = 3(e, f, c\ );\ \ Path_4 = 3(e, f, g).$$

Path scores for these paths; Path 1, 2, 3 and 4 (P_1, P_2, P_3 and P_4 resp.) are

$$P_1 = \frac{Individuals\ in\ a\ Path}{Path^{max}} = \frac{4}{4} = 1;\ P_2, P_3 and\ P_4 = \frac{3}{4} = 0.75$$

Total Score 'T' is

$$T = P_1 + P_2 + P_3 + P_4 = 1 + 0.75 + 0.75 + 0.75 = 3.25$$

Hence the Log Score 'L' can be calculated by

$$L = log_{10}\ T$$

$$L = log_{10} 3.25 = 0.511$$

The log score of the given candidate (CA) is 0.511; but in real life there will be more paths, due to more number of individuals who knew or related to the candidate and more crossovers between the paths than what we depicted in the figure.

Simulation

We also carried out Monte Carlo simulation studies of the above algorithm for a candidate using MATLAB Software. To find the popularity of a candidate we kept overall total number of individuals constant but used varying number of paths and number of individuals in those paths (figure 2). The study has shown that there is a linear relationship between Total Score (T) and the number of paths; that is Total score is approximately equal to the half of the number of paths.

Conclusion:

In this work we have formulated a simple algorithm using which popularity of a candidate (animate or inanimate) can be done. We have introduced a concept of 'Path score' in this algorithm and carried out a Monte Carlo simulation study of the algorithm. Simulation studies had shown a linear relationship between Total Score and number of paths. Further simulation studies and practical experiments using this algorithm were considered by the authors for the future studies.

Reference

1. Antonius HNC and Amanda JR: Understanding Popularity in the Peer System. Curr Dir Psyc Science 14(2),102-105 (2005)
2. Bronwyn EB, Suniya SL: Peer-Perceived Admiration and Social Preference: Contextual Correlates of Positive Peer Regard Among Suburban and Urban Adolescents. J Resrch on Adolescence 17(1),117–144 (2007)
3. Funda M, Elmer VB, Nadeem QM, Kelly KH, Frederick CA, Merrick IR, Henry MK, Raphael EP, Mark AM, Eva SS: Breast cancer on the world wide web: cross sectional survey of quality of information and popularity of websites.BMJ 324(7337), 577–581(2002)
4. Jolene Z, James D: The Faces of Facebookers: Investigating Social Enhancement and Social Compensation Hypotheses; Predicting Facebook™ and Offline Popularity from Sociability and Self-Esteem, and Mapping the Meanings of Popularity with Semantic Networks. J Comp-Medtd Comm 14(1),1–34 (2008)
5. L.A.F. Park and K. Ramamohanarao: Mining web multi-resolution community-based popularity for information retrieval in Proc. CIKM, 545-554 (2007)
6. Lawrence P, Sergey B, Rajeev M, and Terry W: The PageRank Citation Ranking: Bringing Order to the Web. Technical Report. Stanford InfoLab. (1999) http://ilpubs.stanford.edu:8090/422
7. Popularity; www.merrain-websters.com

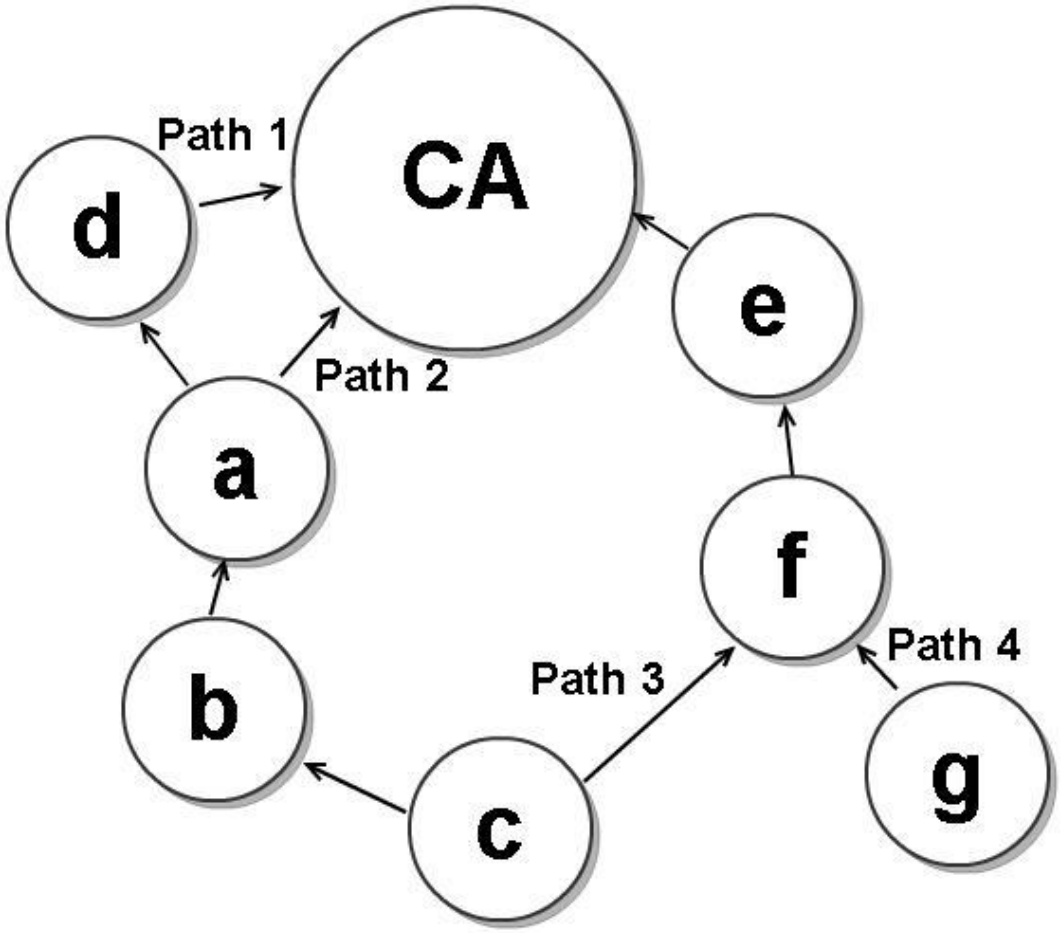

Figure 1: Figure depicting candidate (CA), individuals (a-g) and paths (Path1-Path4)

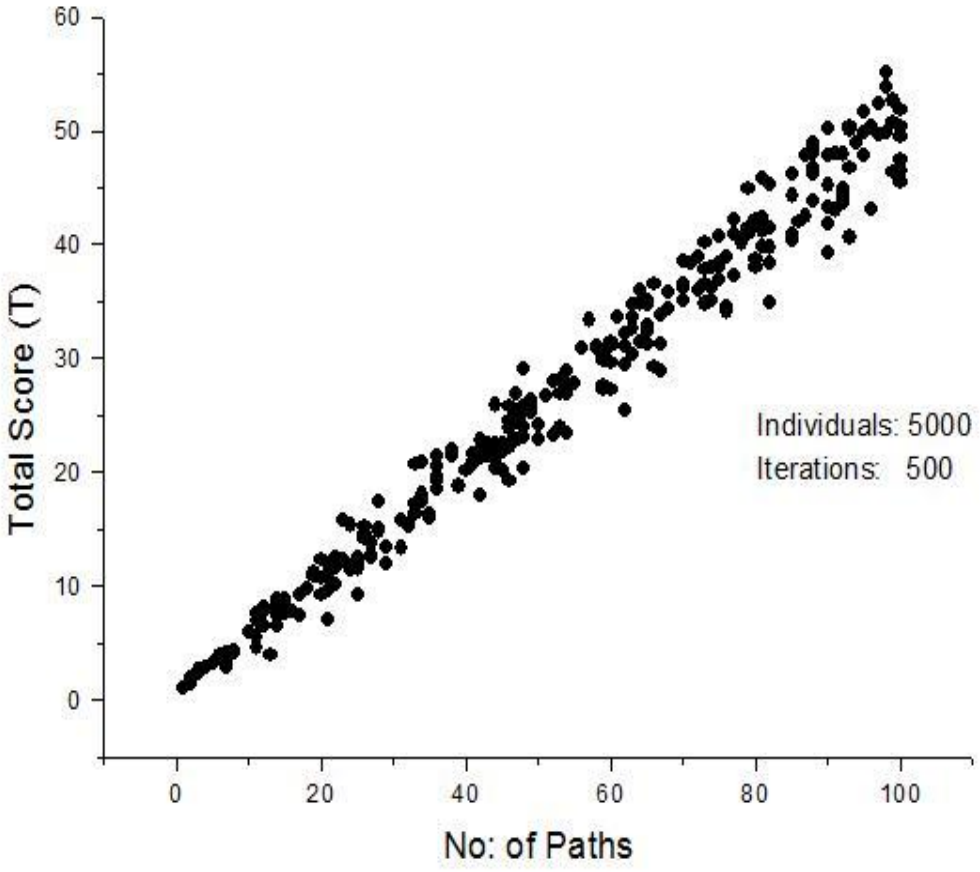

Figure 2: Monte Carlo simulation studies of the algorithm in which 5000 individuals were taken and iterated for 500 times to find total score of a candidate for a given number of paths.