Ranking Designs and Users in Online Social Networks

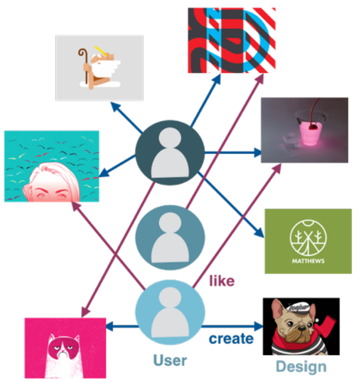


Figure 1. A pictographic representation of a social design network where users share their created designs. Users can also *like* the shared designs.

Biplab Deka
University of Illinois at Urbana Champaign, 201 N Goodwin Ave
Urbana, IL, USA, 61801
deka2@illinois.edu

Haizi Yu
University of Illinois at Urbana Champaign, 201 N Goodwin Ave
Urbana, IL, USA, 61801
haiziyu7@illinois.edu

Devin Ho
University of Illinois at Urbana Champaign, 201 N Goodwin Ave
Urbana, IL, USA, 61801
devinho2@illinois.edu

Zifeng Huang
University of Illinois at Urbana Champaign, 201 N Goodwin Ave
Urbana, IL, USA, 61801
zhuang45@illinois.edu

Jerry O. Talton
Apropose, Inc.
883 N. Shoreline Blvd. #D-220
Mountain View, CA, USA, 94043
jerry@apropose.com

Ranjitha Kumar
University of Illinois at Urbana Champaign, 201 N Goodwin Ave
Urbana, IL, USA, 61801
ranjitha@illinois.edu

CHI'15 Extended Abstracts, Apr 18-23, 2015, Seoul, Republic of Korea
ACM 978-1-4503-3146-3/15/04.
http://dx.doi.org/10.1145/2702613.2732760

Abstract

This work-in-progress presents a new algorithm that leverages social network structure to rank designs and users in online design communities. The algorithm is based on the intuition that the importance of a design should depend on the rank of the users that *created* and *promoted* it, and the importance of a user should depend on the rank of the designs he creates and promotes in turn. The algorithm produces design rankings that are positively correlated with existing social metrics such as number of *likes*, but also allows designs with second-order social import to rise through the ranks. We demonstrate that the algorithm converges, and analyze the rankings it produces on both simulated and scraped social design networks.

Author Keywords

Design; social networks

ACM Classification Keywords

H.3.3 [Information Systems]: INFORMATION STORAGE AND RETRIEVAL – Information Search and Retrieval

Introduction

Online design communities such as Dribbble [2] and Behance [1] allow designers to share their work, get feedback, and find inspiration. These platforms rank designs by social metrics such as the number of *likes*; a

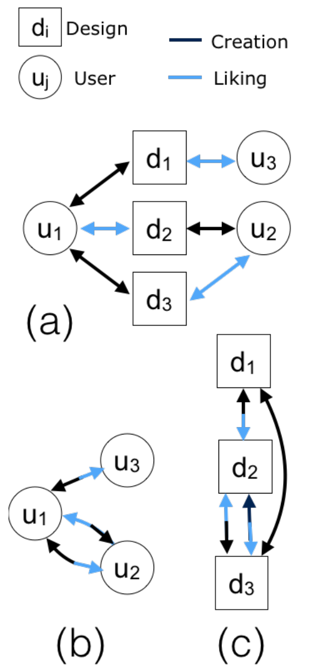


Figure 2. (a) An example *social design network* represented as a graph. (b) The *induced user graph* for the example network. Two users are connected by an edge if they were connected to the same design in the original graph. (c) The *induced design graph* for the example network. Two designs are connected by an edge if they were connected to the same user in the original graph. Both induced graphs may contain up to four different types of edges (Figure 3).

higher ranking leads to more visibility on the site. However, rankings based on these first-order measures ignore both *who* created the designs and *who* promoted them, factors often used to assess the importance of pages on the Web [4], content in social networks [6,7,8], and academic publications [9].

This work-in-progress proposes a design ranking algorithm that leverages the social network structure commonly found in online design communities, where users link to designs through *creation* or *liking* edges (Figure 1). Our approach is motivated by the intuition behind the PageRank algorithm [3] — all citations are not created equal. In our case, not all *likes* are equal: the importance of a design is influenced by the importance of its creators and promoters, and vice versa. For example, a design will be rated higher if it is created or liked by an important user (*i.e.,* famous designer). Similarly, a user will be rated higher if he creates — or likes — important designs.

The algorithm comprises an iterative two-step process that uses design rankings to update user rankings, and user rankings to update design rankings in turn. We verify that the method converges for both simulated graphs and crawled social design networks; a formal proof is future work. The algorithm produces design rankings that are positively correlated with number of likes, but the specific ordering varies significantly from like-based rankings. Future work will examine whether this importance measure leads to improved discoverability of new and better quality designs in online design communities.

Representation

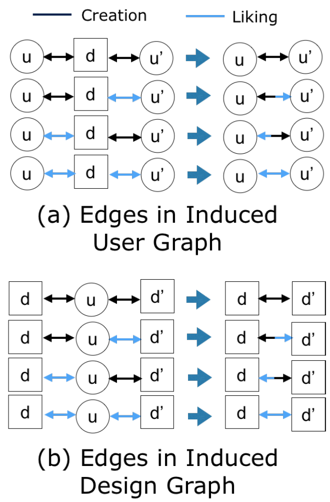
We represent a social design network as a graph $\boldsymbol{G}_{(\boldsymbol{U},\boldsymbol{D})} = (\boldsymbol{U}, \boldsymbol{D}, \boldsymbol{E_l}, \boldsymbol{E_c})$ with two kinds of nodes – *users* and *designs* – and two kinds of edges – *creation* and *likes*. We let $\boldsymbol{U}$ and $\boldsymbol{D}$ represent the set of users and designs, respectively, and $\boldsymbol{E_c}$ and $\boldsymbol{E_l}$ represent the set of creation edges and liking edges, respectively. Since influence is flowing from users to designs in both directions, the edges in this graph are bidirectional.

Given such a network, one popular method for ranking its nodes is the PageRank algorithm [3], which was first proposed for determining the relative importance of webpages. PageRank is based on the assumption that links from important pages should themselves carry more import than links from pages that are relatively unknown. A simplified formula for PageRank is given by

$$PageRank(u) = c \sum_{v \in B(u)} \frac{PageRank(v)}{N_v},$$

where u is a webpage, $B(u)$ represents the pages that link to u, N_v represents the number of outgoing links from v and c is a factor for normalization. Essentially, PageRank computes the stationary distribution of a Markov chain taking a random walk over the network, and assigns each node a rank proportional to the likelihood that the walk visits the node. The PageRank algorithm is numerically stable, and can be computed iteratively and in a distributed manner, making it a useful metric on networks with hundreds of millions of nodes.

Applying PageRank directly to social design networks is complicated by the fact that users and designs — which comprise the nodes of the network — are not directly



(a) Edges in Induced User Graph

(b) Edges in Induced Design Graph

Figure 3. Different types of edges in the original graph (left) induce different types of edges in the induced graphs (right). (a) represents the cases arising in the *induced user graph* and (b) represents the cases arising in the *induced design graph.*

comparable. Therefore, we induce two related graphs of comparable nodes— G_U, a graph of users, and G_D a graph of designs — on which PageRank can be run, and define relationships that allow rank information to be meaningfully transferred between them.

Figure 2 shows the construction of these graphs. Two nodes are connected by an edge in an induced graph if they are both connected to the same node in $G_{(U,D)}$. That is, users are connected to one another by the designs that they have in common, and conversely for designs.

To allow the rankings in the user graph to influence the rankings in the design graph (and vice versa), we weight each edge in one induced graph by the rank of the node generating the edge in the other. For instance, in Figure 2(b), the edge between users u_1 and u_3 is weighted based on the rank of design d_2 because the link $u_1 \leftrightarrow d_2 \leftrightarrow u_3$ exists in the graph in Figure 2(a).

The representation then admits an iterative algorithm, described in detail in the next section: computing PageRank over one induced graph, transferring the calculated weights to the other, and repeating until convergence.

Ranking Algorithm

To understand the ranking algorithm it is helpful to work with matrix representations of the graphs described above. We represent $G_{(U,D)}$ using two matrices – C and L, both of which have as many rows as the number of users in the network and as many columns as number of designs. We call C the *creation matrix* and L the *liking matrix*, where

$$C_{ij} = \begin{cases} 1 \; if \; u_i \rightarrow d_j \in E_c \\ 0 \; if \; u_i \rightarrow d_j \notin E_c \end{cases} \quad \text{and} \quad L_{ij} = \begin{cases} 1 \; if \; u_i \rightarrow d_j \in E_l \\ 0 \; if \; u_i \rightarrow d_j \notin E_l \end{cases}.$$

For example, in Figure 2,

$$C = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad L = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

A non-zero value C_{ij} implies "user i *created* design j". The other direction - "design i *was created by* user j" - is captured by non-zero elements of the transposed matrix C_{ij}^T. Similarly, a non-zero value L_{ij} implies "user i *liked* design j" and a non-zero L_{ij}^T means "design i *was liked by* user j". We also represent the induced user matrix G_U as a square matrix U, and the induced design graph G_D as a square matrix D.

Each edge in the induced user graph G_U is one of four types, given in Figure 3. These four types of edges correspond to the non-zero elements in the matrices CC^T, CL^T, LC^T and LL^T. (To understand why, consider the matrix LL^T. The element in row i and column j in LL^T can be non-zero only if $L_{i,k}$ and $L_{k,j}^T$ are non zero for some k. This means that user i liked design k and user j liked design k as well. This corresponds to the fourth case in Figure 3(a)). This lets us define U as

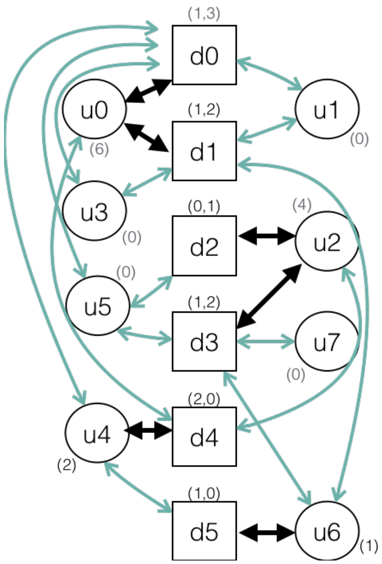


Figure 4. The network used for our evaluations. It consists of eight users and six designs. Four of the users are creators and the other four users are curators (who participate by liking designs). Thick black arrows represent the creation edges whereas the thin green edges represent liking edges. The numbers above each design shows the number of likes separated into two categories – likes from creators and likes from curators. The numbers next to each user shows the total number of likes the designs they created have gotten.

$$U = DiagonalZero(CC^T + CL^T + LC^T + LL^T)$$

where $DiagonalZero(\cdot)$ is a function that sets each diagonal element of the matrix to zero (to avoid self loops in the G_U). Similarly, we can define the matrix D (corresponding to the induced design graph G_D) as:

$$D = DiagonalZero(C^TC + C^TL + L^TC + L^TL)$$

Defined in this way, the non-zero elements in U and D represent the number of edges by which nodes are connected to each other in the respective induced graphs.

This model also allows different types of edges to have different relative importance, for instance to allow the *creation* of a design to factor more prominently in the ranking of a user than the *liking* of a design.

Accordingly, we introduce eight parameters $\{\alpha\}_{i=1}^4$ and $\{\beta\}_{i=1}^4$ to define U and D as:

$$U = DiagonalZero(\alpha_1 CC^T + \alpha_2 CL^T + \alpha_3 LC^T + \alpha_4 LL^T) \text{ and}$$

$$D = DiagonalZero(\beta_1 C^TC + \beta_2 C^TL + \beta_3 L^TC + \beta_4 L^TL)$$

To transfer the computed rankings between induced graphs after each step of the iteration, we assume that user ranks and design ranks are specified as vectors r_u and r_d respectively. Then, U and D are computed as

$$U = DiagonalZero(\alpha_1 C \cdot diag(r_d) \cdot C^T + \alpha_2 C \cdot diag(r_d) \cdot L^T + \alpha_3 L \cdot diag(r_d) \cdot C^T + \alpha_4 L \cdot diag(r_d) \cdot L^T) \text{ and}$$

$$D = DiagonalZero(\beta_1 C^T \cdot diag(r_u) \cdot C + \beta_2 C^T \cdot diag(r_u) \cdot L + \beta_3 L^T \cdot diag(r_u) \cdot C + \beta_4 L^T \cdot diag(r_u) \cdot L)$$

where $diag(r)$ is a diagonal matrix whose diagonal elements are from vector r.

With U and D computed in this manner, we can run PageRank with these two matrices after row normalization. This results in updated user ranks and design ranks which can then be used to compute and update the U and D matrices.

Pseudocode for the method is given in Algorithm 1.

Input: $L, C, \{\alpha_i\}_{i=1}^4, \{\beta_i\}_{i=1}^4$
Output: r_U, r_D
initialize $r_D = (1, \cdots, 1)$;
while *not converged* **do**
  $U = \alpha_1 \cdot C \cdot \mathbf{diag}(r_D) \cdot C^\top$;
  $U \mathrel{+}= \alpha_2 \cdot C \cdot \mathbf{diag}(r_D) \cdot L^\top$;
  $U \mathrel{+}= \alpha_3 \cdot L \cdot \mathbf{diag}(r_D) \cdot C^\top$;
  $U \mathrel{+}= \alpha_4 \cdot L \cdot \mathbf{diag}(r_D) \cdot L^\top$;
  $r_U = PageRank\,(\,(RowNormalize(U))\,)$;

  $D = \beta_1 \cdot C^\top \cdot \mathbf{diag}(r_U) \cdot C$;
  $D \mathrel{+}= \beta_2 \cdot C^\top \cdot \mathbf{diag}(r_U) \cdot L$;
  $D \mathrel{+}= \beta_3 \cdot L^\top \cdot \mathbf{diag}(r_U) \cdot C$;
  $D \mathrel{+}= \beta_4 \cdot L^\top \cdot \mathbf{diag}(r_U) \cdot L$;
  $r_D = PageRank\,(\,(RowNormalize(D))\,)$;
end

Algorithm 1. The algorithm for computing design and user rankings.

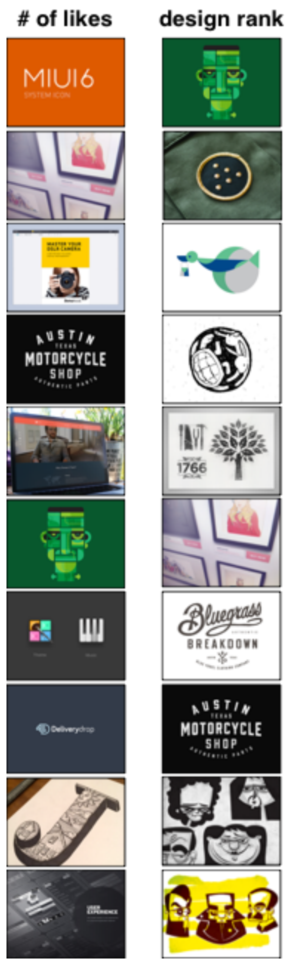


Figure 5. The ten highest ranked designs (from top to bottom) in our real world dataset, using number of likes (left) and our algorithm (right).

Results

Simulated Network

We ran our ranking algorithm on a randomly generated network comprised of eight users and six designs (Figure 5), using two different sets of edge parameters. In both cases, the design and user rankings converged and are consistent with the assumptions made about the network. We compare our results to like-based ranking schemes: ranking designs by their number of likes, and users by the number of likes their designs have received (Table 2).

For the first experiment, we simulated a network where rank flows in only one direction: users only receive rank from their creations, and designs only receive rank from their promoters. These assumptions are encoded in our model by setting the parameter values $\alpha_1, \beta_1 = 1$ and $\alpha_2, \alpha_3, \alpha_4, \beta_2, \beta_3, \beta_4 = 0$.

The results of the ranking algorithm are shown in Table 2. We observe that the highest ranked design is d_4, which can be explained by the fact that it has two likes from two creators. This results in d_4's creator u_4 becoming the highest ranked user. This in turn gives the design he liked (d_5) a high rank, and so on. We also note that users that created no designs (*i.e.,* curators) have the lowest rank since there is no rank that flows to the curators from other users.

Not all curators should have the same rank: an online community should incentivize curators to help identify good designs. Therefore, in a second experiment we allow users to receive rank from designs they create *and* like, but rank from a created design is weighted ten times more than rank from a liked design.

Similarly, in the second experiment, we allow designs to receive rank from *both* their promoters and their creators, but rank from a promoter is still weighted ten times more than rank from a creator. We encoded these assumptions in our model by setting parameter values $\alpha_1, \beta_1 = 10$, $\alpha_4, \beta_4 = 1$ and $\alpha_2, \alpha_3, \beta_2, \beta_3 = 0$.

Running our algorithm with this new set of parameter values, we observe that overall the ranks are similar to the last experiment but now d_1 ranks higher than d_3 since d_1's creator (u_0) is ranked higher than d_3's creator (u_2). Also, the curators now are able to have different ranks based on the rank of the designs they liked.

	Ranks	# of likes	Exp. 1	Exp. 2
Designs	1	d0	d4	d4
	2	d1, d3	d5, d0	d5
	3	d4	d1, d3	d0
	4	d5, d2	d2	d1
	5			d3
	6			d2
Users	1	u0	u4	u4
	2	u2	u0	u0
	3	u4	u6	u6
	4	u6	u2	u2
	5	u1,u3,u5,u7	u7,u5,u3,u1	u5
	6			u3,u1
	7			u7

Table 2. Results from running the ranking algorithm over the simulated social network in Figure 4 using two different sets of edge parameters.

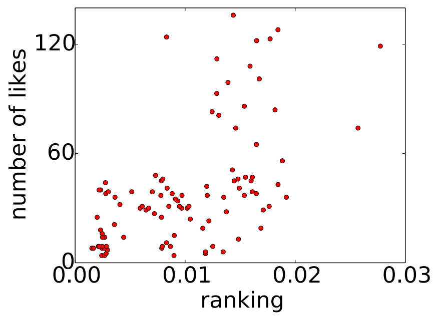


Figure 6. The design rankings produced by the algorithm are positively correlated with the number of likes.

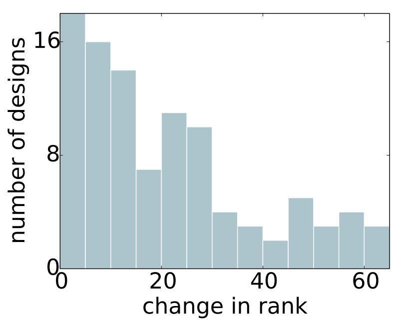


Figure 7. Histogram of the change in rank of individual designs between our rankings and like-based rankings.

Real World Network
We also tested the ranking algorithm on a graph sampled from Dribbble [1], an online social design network. The network we sampled consisted of 100 designs and 2592 users; 68 users had *created* at least one design and the rest had *liked* at least one design. The number of likes per design ranged from 4 to 136 ($\mu \approx 40.1, \sigma \approx 32.5$), and the network contained 4009 total likes.

We ran the ranking algorithm over the network using the parameter settings from the second experiment ($\alpha_1, \beta_1 = 10$, $\alpha_4, \beta_4 = 1$ and $\alpha_2, \alpha_3, \beta_2, \beta_3 = 0$), and the design and user rankings converged. The design rankings produced by the algorithm are positively correlated with the number of likes (Figure 6). However, the specific ordering varies significantly from like-based rankings: an individual design shifted 21 spots in the ranking on average (Figure 7).

Figure 5 shows the top ten ranked designs based on number of likes and our ranking algorithm. We observe that only three out of the ten designs are common in both rankings. The tenth design in the 'design rank' column only had 29 likes, but was still able to break into the top ten based on secondary network effects. Future work will examine if this importance measure is a better predictor of quality than number of likes.

Discussion and Future Work

This work-in-progress presents a novel algorithm for ranking designs and users in a social design network. Early results demonstrate that the generated rankings depend on how different types of edges are weighted. In the future, we hope to learn these parameter values from data to produce rankings that lead to improved discoverability of new and better quality designs in online communities.

Moreover, social networks have many other types of social features [5]: users can *follow* other users, designs can be grouped together into *collections*, etc. How can we augment the model to capture these additional factors that could affect rank?

Finally, these importance measures could conceivably help identify and reward different types of users in online communities. The top *curator* isn't necessarily the top *designer*; nonetheless, curators play an important role in creating the network structure that keeps online communities engaging.

References

[1] Dribbble dribbble.com

[2] Behance behance.net

[3] The PageRank Citation Ranking: Bringing Order to the Web Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd *Stanford InfoLab Tech Report 1999*

[4] Easley, David, and Jon Kleinberg. "Networks, Crowds, and Markets."*Cambridge University*.

[5] Gilbert, Eric, et al. "I Need to Try This!: A Statistical Overview of Pinterest." *Proc. CHI*, 2013.

[6] Halberg, Victor et al. "Tweetrank: An adaptation of the PageRank algorithm to the Twitter world." *Tech Report*

[7] Kwak, Haewoon, et al. "What is Twitter, a social network or a news media?."*Proc. WWW*, 2010.

[8] Ravikumar, Srihith et al. "Ranking tweets considering trust and relevance." *Proc. IIWeb,* 2012.

[9] Yizhou Sun, et al. "Ranking-based clustering of heterogeneous information networks with star network schema." *Proc. KDD,* 2009.