LEARNING DESIGN TRENDS FROM SOCIAL NETWORKS

Data Mining, Analysis & Visualization of Grasshopper® Online User Community

IMMANUEL KOH
1 École polytechnique fédérale de Lausanne (EPFL)
1 immanuel.koh@epfl.ch

Abstract. The paper has demonstrated that the increasingly online relationship between designers and their digital tools can be quantitatively represented, described and analyzed through the data-mining of design-domain specific and tool-based social network (i.e. Grasshopper3D). It explores design trends’ correlations based on network user groups’ size, users’ demographics, nodes’ degree centrality and discussion threads’ popularity.

Keywords. Social Networks; Design Trends; Big Data; Parametric Design Tools; Data Visualization.

1. INTRODUCTION

Data mining and analysis of social networks has become indispensable to numerous retail companies today, in extracting invaluable insights to consumers’ demographics, online behaviours and potential product preferences. This ubiquitous use of online social media and its consequent abundance of data produced, has also made the learning and prediction of near future trends a feasible endeavour for researchers today. Today’s designers and architects are increasingly using social media platforms to search for inspirations, discuss state-of-the-art design techniques, showcase their design portfolios, as well as, to craft their online design profiles. New open-source models of design data within architecture (Ratti and Claudel 2015) and current formal tendencies of Parametricism (Schumacher 2009) has also suggested the significant role of social media networks in altering collaborative curation, discussion and digital design among today’s architects. Traditionally, architectural design discourse was mainly limited to offline activities, such as exhibitions, book publications and design reviews, where the curator or author would steer the design trend and its audience from a hierarchically privileged position. More bottom-up and dynamic nature of online communication and trending facilitated by today’s social media is quickly beginning to challenge this earlier model.

Although popular social media platforms, such as Facebook and Twitter, have already become ubiquitous to our daily digital life; their predominantly social purpose tend to be too generic for a typically task-driven designer users

base. Their limited expressivity of content formats (e.g. Twitter’s original 140 max. characters and Facebook’s Likes) is equally inadequate for the needs of discourses in design. Such atypical needs might include the sharing of specific formats of digital design files, extended discussions of new design techniques, experimentations with new software’s features and analysis of both geometrical and mathematical concepts. Apart from the expressivity of permitted content on the network, the structure of the network itself needs to be appropriated for the ways in which designers carry out social networking. Research in design cognition can be found in psychology (Tversky and Hard 2009) but none has approach the topic from a large scale and data-driven manner, beyond a handful of interview subjects limited in an academic laboratory context. Quantitative research in digital humanities have typically focused on more generic social media platforms and but rarely on both domain specific and task-oriented user base (including designers).

The research here aims to data mine, analyse and visualise design-related social networks, to gain an understanding of the current state of design creations in architecture, while anticipating near future design trends by evaluating their various potential correlations. The focus of the paper is to describe its findings, rather than to provide a comprehensive explanation of it. This paper will use a case study (Grasshopper® online user community) to answer the 4 research questions (R1-R4). Grasshopper/GH is the fastest growing digital design tool platform used today for computational and parametric architectural design (Day 2009). Founded in 2007, it now has a user base of more than 45,000 (as of 24th June 2016). More specifically, its vibrant online social network provides the ideal users’ dataset for our research hypotheses. Unlike the relatively larger numbers of registered members on the website, there are only 4616 members on its Facebook page & 155 followers on its Twitter page. (as of 24th June 2016). Thus, the logical decision is to use the user’s data available on the website to carry out the research. Although one might argue that it is more of an internet forum than a social network, the position of the paper is that it is a loose combination of both. Strictly speaking, traditional forum revolves around topics (i.e. ‘bulletin-board’), while typical social network revolves around persons. In that sense, the Grasshopper® website is on the one hand, a forum that provides the means to post and respond to specific discussion topics/threads, as well as; on the other hand, a social network that provides the means to ‘share’ or ‘like’ the works of a member, to ‘befriend’ him/her or to ‘join’ his/her affiliation via group memberships. The data extraction process was carried out over a duration of 5 days with custom-written web-crawlers. The results were further processed for the final network analysis and visualisation. The 4 research questions are as follow:

- **R1- Design trends as a function of algorithmic tools’ user group size:** Which are the most popular GH add-ons and what are the sizes of these user groups in the social network? Is there a correlation between design trends and the popularity of specific algorithms?
- **R2- Design trends as a function of demographics:** What is the typical membership profile, in terms of their nationality, age and gender? Is there a correlation between design trend and the socio-cultural background of social network members?
• R3- Design trends as a function of network’s node degree centrality: Who are the most connected and influential users in the social network? Is there a correlation between design trend and the idiosyncratic design preferences of a network’s supernodes?

• R4- Design trends as a function of most prevailing forum discussion thread: What is currently the most talked about design topics in the social network? Is there a correlation between design trend and the most prevailing design topic?

2. DATA & METHODS

There is neither an official or open-source API nor database available to access the data on the website. As a result, a number of web-crawlers have been written to scrap the necessary data from the Grasshopper3D (GH) website. The web-crawlers begin by first traversing the ‘Members’ page, the ‘Add-ons’ page and the ‘General Discussion’ page found on the respective hyperlinks on the homepage’s main menu bar. The ‘Members’ page consists of a list showing a paginated view of all the individual registered member, the ‘Add-ons’ page consists of a list showing a paginated view of all the registered groups and the ‘General Discussion’ page consists of a list showing a paginated view of all the discussion topics posted by the members. The web-crawlers would then traverse each of the member profile page, group profile page and discussion page. From the user’s profile page, the user’s basic data can be extracted, such as profile image, name, gender, city, country, likes on Ning and Facebook, tweets, blog posts, events, photos, photo albums and videos. However, more critical information needed for the research are the hyperlinks to the pages showing the particular user’s own friends, groups joined and discussion posted. From the group page, all replies posted which are tagged with the name of the group can be extracted, in addition to the basic description of the group itself. From the discussion page, all posts with or without tags could be extracted. These extracted data are then processed as nodes and edges of the respective networks for further analysis and visualization. These relationships concerned are namely: Users - Group, Groups - User, Users - Country, Users - User, Replies - Group, Replies - User and Replies - Topic. The network visualizations are made using the open-source tool called Gephi with the data input formatted as csv and implemented with the Force-Atlas 2 layout algorithm, while the web-crawlers code is written in Python using Scrapy library. There may be privacy concerns regarding the exposure of specific GH usernames in the paper and how one should address them accordingly. The author’s position is that since all scrapped data presented here is already visible and accessible in the public domain, it is the relationships that are being revealed, not the identities per se.

3. RESULTS

The results are visualized, analyzed and discussed in the following categories, relating to the 4 research questions stated earlier in the paper:
3.1. R1. DESIGN TRENDS AS A FUNCTION OF ALGORITHMIC TOOLS’ USER GROUP SIZE

**Users/Group** (see Figure 1): There are 223 groups (i.e. ‘Add-ons’) registered with 16% of all users (i.e. 7147 out of 45,192). This network consists of a directed graph with 7370 nodes (2 types: users & groups) and 23,986 edges (1 type: between users & groups). In this ‘users-group’ visualization, the size of each group node (incl. annotations) is shown proportional to the number of members connected to it. The number of users per group has a maximum of 2106, minimum of 1, mean of 108 and median of 36.5. The 1st Ranked Group (*Kangaroo*) has 30% of the total users (i.e. those who have joined >=1 group). Top ranked groups are mostly clustered near one another, suggesting the groups tend to share similar users. The network distribution suggests that complementary tools tends to cluster. For example, *Galapagos* (Evolutionary Solver Tool) both complement *Kangaroo* (Physics Engine Tool) and *GeometryGym* (Building Information Modeling Tool) as an optimization tool. Educational or Social-based groups typically has lesser users. For example, the largest social group is *London-UK GH Group* (166 members), while the largest educational group is *Parametric Primer* (265 members). It is important to note that Tool-based groups attract most users and are the only ones capable of having thousands of users. The result of the research shows that the popularity of a GH add-on and the size of its user group is dependent on the tool’s ability to complements other significant GH add-ons, as well as, other non-Rhino software packages. There is a correlation between the current rise of BIM in design production and the most popular GH add-ons found in the analysis.

![Network showing no. of user nodes per group node.](image)

**Groups/User** (see Figure 2): This ‘groups-user’ visualization shows the size of each user node (incl. annotations) as proportional to the number of groups he/she is connected to. The number of groups per user has a maximum of 84, minimum of
1, mean of 3.35 and median of 1. Accordingly, a user joining more groups does not seem to correlate with more activities in either the discussion group or the number of friends. For example, it can be noted that the 1st Ranked LuisGarciaLara (with 26 discussion posts & 21 friends), 3rd Ranked GiulioPiacentino (with 1554 discussion posts & 587 friends) and 4th Ranked AndrewReynolds (with 0 discussion post & 0 friends) differ greatly in an inconsistent manner. However, top ranked users can be seen mostly clustered near one another, suggesting the tendencies for them to share similar groups.

![Figure 2. Network showing no. of group nodes per user node.](image)

3.2. R2. DESIGN TRENDS AS A FUNCTION OF DEMOGRAPHICS

**Users/Country** (see Figure 3): There are 45,192 Designers (i.e. ‘Members’) and 301 corresponding countries registered on their profile pages. This network visualization consists of a directed graph with 45,493 nodes (2 types: users & countries) and 45,026 edges (1 type: between users & countries). The number of users per country has a maximum of 10,796, minimum of 1, mean of 147.6 and median of 4. This ‘users-countries’ visualization shows the size of each country node (incl. annotations) as proportional to the number of users connected to it. Users from USA disproportionately (24%) dominated the total number of users. This 1st ranked country (USA) is also 4 times the size of the 2nd ranked country (UK). In fact, the top 3 English-speaking countries constitute 1/3 of the entire user base. Originally, other desired data included the gender of each user; however, it was eventually forsaken due to an incomplete dataset when the website made such
declaration non-compulsory a few years ago. The result of the research shows that there is a disproportionate domination of users from English-speaking countries. There is a correlation between the greater use of parametric design tools in these countries and the most dominant nationalities found in the analysis.

Figure 3. Network showing no. of user nodes per country node.

3.3. R3. DESIGN TRENDS AS A FUNCTION OF NETWORK’S NODE DEGREE CENTRALITY

Users/User (see Figure 4): There are 10,794 users who have befriended more than 1 other user, thus only representing 24% of all users (i.e. 10,794 out of 45,192). This network consists of an undirected graph with 10,794 nodes (1 type: users) and 30,215 edges (1 type: between users & users). The number of friends per user has a maximum of 1661, minimum of 1, mean of 5.6 and median of 1. This ‘users-users’ visualization shows the size of each user node (incl. annotations) as proportional to the number of friends (i.e. other users) connected to him/her. 3 users (ranked 1st-3rd) disproportionately dominated the total number of users and each knows approx. 15% of entire network, while having separate communities of friends along the fringe of the entire network. 2 (ranked 4th & 5th) of the top 5 users are architects, but each knows only approx. 5% of entire network, while having a more evenly distributed in the middle of the entire network. Who are the top 3 most popular users? They are AndresGonzalez (1st) who is the RhinoFabStudio world-wide director with a background in Mechanical Engineering, JonMirtschin (2nd) who is the director at Geometry Gym (Rhino Plugin) with a background in
Structural Engineering and Bob McNeel (3rd) who is the CEO at Robert McNeel & Assoc (Rhino) with a background in Business Finance. It takes a maximum of 12 friends to have the entire network to be connected. There are 119 communities with 26 of these being most significant in numbers. There is also no significant difference in results using either Node Degrees or Eigenvector Centrality. The network has a diameter of 12, modularity of 0.451, 119 communities and an average path length of 3.5. The result of the research shows that the most connected users are those who play the expert technical role of solving questions posted by other users relating to parametric design problems. Thus, these supernodes tend to be GH add-ons developers who either work at commercial/academic research groups or at/with Robert McNeel & Assoc (Rhino) itself. There is no clear idiosyncratic design preference found in these supernodes that might suggest an obvious influence on design trend, although their recommendation of parametric references in their posts might indirectly influence others’ aesthetic preferences.

3.4. R4. DESIGN TRENDS AS A FUNCTION OF MOST PREVAILING FORUM DISCUSSION THREAD

Replies/Group (see Figure 5): There are 128 groups (i.e. ‘Add-ons’) tagged with 18% of all discussion replies (i.e. 4029 out of 22,325). This network consists of a directed graph with 4157 nodes (2 types: replies & groups) and 4,029 edges. The number of replies per group has a maximum of 1236, minimum of 0, mean of 31.5 and median of 5. This ‘replies-groups’ visualization shows the size of each group.
node (incl. annotations) as proportional to the number of reply nodes connected to it. Unlike the earlier statistics showing number of users per group, it seemed a more accurate way of measuring the significance of the Group/Tool based on the number of tagged replies.

Replies/User (see Figure 6): There are 18.5% of all users who replied (i.e. 8,332 out of 45,192), while there is a total of 22,325 replies. This network consists of a directed graph with 30,657 nodes (2 types: users & replies) and 102,132 edges. The number of replies per user has a maximum of 11,961, minimum of 1, mean of 12.51 and median of 2. This ‘replies-users’ visualization shows the size of each user node (incl. annotations) as proportional to the number of reply nodes connected to it. The 1st Ranked User (David Rutten) is the creator of Grasshopper3D (GH) and disproportionately dominate the entire discourse with 54% of all replies. It is also more than twice the number of replies compared to the 2nd Ranked User. There exists a small cluster of users (see bottom of network viz) that mainly replies to similar topics also replied by 5th Ranked djordie and 11th ranked HannesLoeschke. This suggests sub-discourses among designers creating sub-communities as a result.
Replies/Topic (see Figure 7): The number of replies per topic has a maximum of 245, minimum of 1, mean of 4.57 and median of 3. This ‘replies-topics’ visualization shows the size of each topic node (incl. annotations) as proportional to the number of reply nodes connected to it. Almost half of the Top 25 Ranked Topics pertains to new releases of Grasshopper versions. For example, grasshopper-0-8-0051-available-for-download...etc. Other Top Ranked Topics include the development of the software itself. For example, my-ideas-for-future-features (Ranked 16), continued-developer-absence (Ranked 18). The community’s response to congratulate-david-rutten-on-his-acadia-2012-award (Ranked 12) shows the immense popularity of the software and its creator (David Rutten). The remaining Top Ranked Topics consists of interesting and even passionate discussion/tutoring on specific aspects of design. For example, how to generate a 3D mesh in skeletal-mesh (Ranked 12) and how to use fibonacci logic to generate designs in fibonacci...etc. The result of the research shows that, in addition to the mostly technical discussion online, there are extended discussion threads which are inspirational and motivates participation by other users in exploring certain algorithmic aesthetics in greater depth. There is, however, no strong correlation to design trend and explicit discussion of design topics found.
4. CONCLUSION

This paper has explored the dynamics of a design-domain specific and tool-based social network (i.e. Grasshopper3D), which is characterized by a highly technical and task-driven perspective. The paper has also demonstrated that the increasingly interwoven relationship between designers and their digital tools can be quantitatively analyzed as a graph-based dataset in gaining some insights on the changing landscape of architectural design trends. However, the author is aware that there is a lack of any discussion on how these insights could then be directly translated and applied in architectural design itself. We hope that this early attempt will at least serve as a plausible research framework for uncovering implicit correlations between the nature of social networks and our near future architectural design trends in a data-driven manner. Future research development could include a timeline functionality to visualize the evolution of the designer network. In addition, texts (NLP), images (CV), videos and Rhino GH files could be extracted and processed for further analysis. The visualization could also be made online to receive feedbacks from GH users themselves.

References


