Remix in 3D printing:
What your sources say about you

Spiros Papadimitriou[‡] Evangelos Papalexakis[§] Bin Liu[‡] Hui Xiong[‡]

[‡] Rutgers University
New Brunswick, NJ, USA
{s.papadim,binben.liu,hxiong}@rutgers.edu

[§] Carnegie Mellon University
Pittsburgh, PA, USA
epapalex@cs.cmu.edu

ABSTRACT

Concurrently with the recent, rapid adoption of 3D printing technologies, online sharing of 3D-printable designs is growing equally rapidly, even though it has received far less attention. We study *remix* relationships on Thingiverse, the dominant online repository and social network for 3D printing. We collected data of designs published over five years, and we find that remix ties exhibit both homophily and inverse-homophily across numerous key metrics, which is stronger compared to other kinds of social and content links. This may have implications on graph prediction tasks, as well as on the design of 3D-printable content repositories.

1 INTRODUCTION

Rapid prototyping and manufacturing technologies, especially 3D printing, have grown explosively in the past few years, substantially lowering barriers to entry in designing and making physical objects. Objects can be rapidly and easily manipulated in a computer, and can be published and shared [2, 3] (similar to, e.g., documents, software, or music). As a result, online sharing of 3D-printable designs is growing rapidly [10]. This can have implications on the design and innovation processes for physical products and, therefore, studying it is important. A key socio-technical phenomenon is *remixing*, where one or more *source* designs are combined and modified, to produce a *derivative* design.

We collected real data spanning several years from from Thingiverse, the most popular online content publishing and social network platform for 3D printable designs. The extracted data can be naturally modeled as a collection of graphs with numerical node attributes.

Our main contributions are: *(1) Methodology*: We propose a graph auto-regressive model, suitable for analyzing remix networks with node attributes; *(2) Experiments*: Our analysis is based on real data we collected, consisting of designs published over a span of almost five years; *(3) Observations*: To the best of our knowledge, we are the first to observe homophily and inverse-homophily in remix networks. We frame the question of remix link value as a graph auto-regression problem, extending widely used measures of assortativity [9].

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org//10.1145/2740908.2745943.

2 DATA

The data consist of all 36,504 public "things" published from September 2008 to March 2013 on Thingiverse, 8,126 users, and 4,373 collections (average size 11.5 items). The two main entities are *things* and *users*. Things are *created* by exactly one user.

A relation of particular interest, and the central topic of this work, is a *remix*. Creators can indicate that their design "remixes" another. This is a directed many-to-many relationship, from a *source* thing which is remixed into a *derived* thing. A remix relation indicates *creative affinity* and it is a form of self-declared credit, where a remix: (i) is a direct derivative of the source, often re-using design files; (ii) is inspired by the source thing, even if it is reworked from scratch or a complete re-purposing of the original; (iii) uses or extends the source thing (e.g., add-on parts, or component re-use). Any user can also curate named *collections* and add things to them. The number of collections in common between two things is another good proxy for design affinity.

The relationships of interest among designs are naturally formalized as a graph $G = (V, E)$. Nodes $t \in V$ always correspond to *things*, and have a set of N numerical attributes, $x_i(t)$, $1 \leq i \leq N$. We compare three graphs, which capture associated but different relationships between things. The first is the *remix graph* G_R, where edge $(t_s, t_d) \in E_R$ is present if t_s is one of the source designs for derivative t_d (see `http://bitquill.net/make/remix/` for interactive visualization). Similar to G_R, we define the co-collection and same-author graphs, G_C and G_A, where an edge is present if two things appear together in the same collection(s), or are created by the same author, respectively.

3 ASSORTATIVITY ANALYSIS

The rich combination of attributes (user interest, popularity, and content), and relationships available on Thingiverse allows us to compare the value of these relationships as a regression problem.

In a preliminary analysis [10], we performed standard multivariate regression across thing attributes, ignoring graph structures. Following a widely used measure of network homophily [9], we extend our analysis, proposing an auto-regressive graph model, which assumes that the value $x_i(t)$ of the i-th attribute of node t depends both on the values $x_j(t)$ of the other attributes of the same node, as well as the values $x_j(t')$ of it's neighbors $(t', t) \in E$. Letting $\tilde{m}_i^2 :=$

$\sum_{t\in V} x_i(t)/|V|$, $\tilde{\mu}_i^2 := \sum_{t\in V} |\mathcal{N}'(t)|x_i(t)/|E|$, $\tilde{s}_i^2 := \sum_{t\in V} x_i^2(t)/|V| - \tilde{m}_i^2$, and $\tilde{\sigma}_i^2 := \sum_{t\in V} |\mathcal{N}'(t)|x_i^2(t)/|E| - \tilde{\mu}_i^2$, estimate covariances $\tilde{r}_{ij}^2 := \left(1/|V| \sum_{t\in V} x_i(t)x_j(t) - \tilde{m}_i\tilde{m}_j\right)/\tilde{s}_i\tilde{s}_j$, and $\tilde{\rho}_{ij}^2 := \left(1/|E| \sum_{(t,t')\in E} x_i(t)x_j(t') - \tilde{\mu}_i\tilde{\mu}_j\right)/\tilde{\sigma}_i\tilde{\sigma}_j$, which assign equal weight to each node, and to each edge [9], respectively. We then estimate t-test values (adjusting sample weights as necessary), to measure statistical significance.

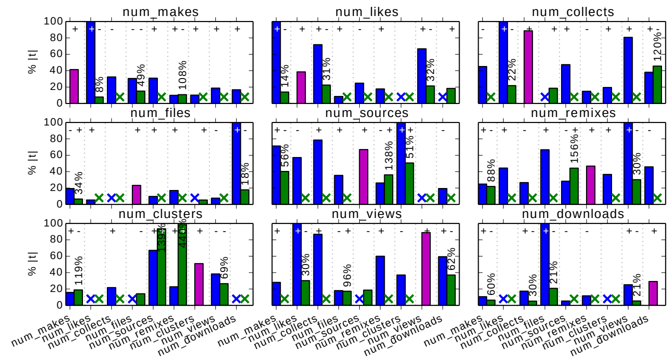
Results. We use two kinds of thing attributes: (i) inherent metrics of importance or popularity (number of views, downloads, likes, makes, and collections containing the thing) or design complexity (number of files), and (ii) structural metrics including in- and out-degree (number of sources and of remixes) and clustering coefficient. Most features represent highly skewed counts, hence we apply a logarithmic transformation (effectively fitting log-linear models). We apply a logit transformation to the the local clustering coefficient.

Figure 1 plots the normalized *absolute* t-score values (higher is better) for both self-node (blue) and cross-neighbor (green and purple) features. Green bars are additionally annotated with their value relative to the corresponding blue bar. We mark features having $p > 0.01$ with an $\times$. The sign of the corresponding regression coefficients is indicated above each bar. t-scores are related across networks but not identical, since the samples (nodes) may differ (we remove disconnected nodes). Surprisingly, the co-collection network (omitted for space) carries little information, with only a few significant correlations, which arise effectively by definition.

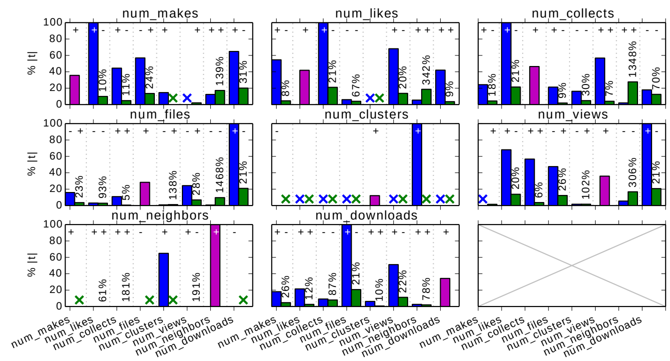
Compared to other networks, neighbor features have the highest overall significance in the remix network. Starting with same-attribute homophily, we observe homophily on the number of files; complexity generally stays the same (the stronger significance of downloads is likely due to how those are counted). Next, the number of views on neighbors (remix sources) is among the best predictors for the number of views on the node itself. The same is true for number of collections, possibly because designs related via a remix are often viewed together, or collected together for future reference. On the other hand, the same is not true for the number of downloads and makes, likely because users choose only one of those related designs. Moving on to cross-neighbor features, the number of remixes of neighbors is almost as good a predictor of the number of makes as is the number of remixes of the node itself (108%), but with an interesting twist: the coefficient for neighbor remixes is negative, whereas for self remixes it is positive. The same is true for the reciprocal relationship.

More often than not, the sign of the coefficients for self and neighbor features are opposite, although not always in the same direction. Notably, the effect of the number of sources on the number of (re-)remixes is negative for the self-feature, but positive for the neighbor feature. Exceptions include the effect of the number of files on downloads. As previously observed [10], these have a negative effect on downloads (people prefer simpler designs) and the same is true for neighbors' files (complex sources have a further negative effect on downloads).

Related work. To the best of our knowledge, very few data-driven studies analyze 3D-printable digital content [10, 7], and try to understand the technology's broader implications. Remixes have been studied in other domains, including [11, 4, 8, 6]. Measures of assortativity were introduced by [9], and studied further, e.g., [1, 5]. We build upon some of that work.



(a) Remix network.



(b) Same-author network.

Figure 1: Predictive power of self and neighbor features.

4 CONCLUSION

3D printing turns physical things into digital content, bringing the power of the web to their design and production. Based on data we collected from Thingiverse, which combines social and content networks with inherent attributes of design importance, popularity, and complexity, we are able to approach the question of the value of remix links as a regression problem, by extending standard measures of assortativity into a graph auto-regression model, allowing us to gain insights about the usefulness of the remix mechanism.

REFERENCES

[1] A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *KDD*, 2008.

[2] C. Anderson. *Makers: The New Industrial Revolution*. Crown, 2012.

[3] Y. Benkler. *The Wealth of Networks*. Yale Press, 2006.

[4] G. Cheliotis and J. Yaw. An analysis of the social structure of remix culture. In *C&T*, 2009.

[5] T. L. Fond and J. Neville. Randomization tests for distinguishing social influence and homophily effects. In *WWW*, 2010.

[6] B. M. Hill and A. Monroy-Hernandéz. The cost of collaboration for code and art: Evidence from a remixing community. In *CSCW*, 2013.

[7] H. Kyriakou and J. Nickerson. Idea inheritance, originality, and collective innovation. In *WIN*, 2013.

[8] K. Luther, N. Diakopoulos, and A. Bruckman. Edits & credits: Exploring integration and attribution in online creative collaboration. In *CHI*, 2010.

[9] M. E. Newman. Mixing patterns in networks. *Phys. Rev. E*, 67, 2003.

[10] S. Papadimitriou and E. Papalexakis. Towards laws of the 3D-printable design web. In *WebSci*, 2014.

[11] R. Shaw and P. Schmitz. Community annotation and remix: a research platform and pilot deployment. In *HCM*, 2006.