RESEARCH NOTE

Remix culture: an empirical analysis of creative reuse and the licensing of digital media in online communities

Giorgos Cheliotis (PI)
School of Information Systems
Singapore Management University
80 Stamford Road
Singapore 178902
giorgos@smu.edu.sg

Contributors

Clint Mark Lumantao Gono, Ankit Guglani, Mike Linksvayer, Victor Stone, Warren Chik, and Giri Kumar Tayi

January 10, 2007

This is a short description of work in progress, to communicate preliminary findings and to propose more research in this direction. Results will likely change as we continue to refine and extend the analysis.

Abstract

We explore the nature and impact of creative reuse in the production of digital media by analyzing the output of an innovative online music sharing community, *ccMixter*. The dataset is of great significance because this is one of only a handful of online communities which do not only allow for the sharing of user-generated content, but also track the evolution of content *after* it has been published online and encourage reuse of the content for the production of new works. All content on *ccMixter* is legally uploaded, copyrighted, and licensed under Creative Commons. Much has been written about the birth of a new "remix culture" on the Internet and how collaborative Web 2.0 technology has led to an explosion of user-generated content. But very little is known about the process of developing digital media in an open and collaborative fashion, the incentives of participating authors, and the outcomes of their actions. Based on our earlier studies of Creative Commons licensing [1] and the analysis of this unique online community we hope to shed more light on the structure and dynamics of such activities by providing some of the first visualizations ever of large-scale remixing activity and presenting our preliminary findings.

Introduction

Creative reuse is the process whereby one or multiple works, or parts thereof, are combined into a new work that is original, i.e. a non-obvious extension, interpretation or transformation of the source material. The practice of reuse is widespread in our society and permeates many otherwise unrelated activities, from industrial manufacturing (building complex systems out of simple multi-purpose parts) to software design (code reuse), and from scientific publishing (reuse and citation of prior work) to fashion design (reuse of patterns, fabrics and designs), just to name some examples.

While the creative process is still to some extent associated with the archetypal image of a single creator working copiously and in isolation to produce a work that is entirely personal, in the age of the participative Web new works are increasingly the result of collaborative and iterative efforts between small or even very large groups of user-creators, utilizing shared pools of reusable and malleable source material. This evolution towards more social modes of cultural production exemplifies and amplifies the significance of the observation that "creativity always builds upon the past", a phrase commonly attributed to Professor Lawrence Lessig, founder of Creative Commons [2]. Creative Commons (CC) is a set of licenses with which authors can grant their audience certain freedoms (such as the freedom to remix a work), thus voluntarily relaxing some of the restrictions of copyright.

Since the launch of Creative Commons in late 2002, the licenses and the communities that use them have grown to millions of users and have spread around the world, as our earlier research has shown. Yet still today very little is known about the incentives of authors who will permit the creation of adaptations of their works by the general public. Copyright law forbids the creation of such derivative works without the explicit permission of the author, irrespective of whether these are produced for profit or not. This is done to protect the author's interests, but in the face of widespread online copying and remixing of works (with or without author consent), and with significant recent indications that the music industry is gradually moving away from technical protection measures[1], we must ask these questions:

- How can we protect authors' interests in the age of participation and how can we optimally balance private interests against the interests of a society that is increasingly depending on open access to information?
- What happens to all the content that is copied and remixed online? What is the value-add, if any, in cultural or in economic terms, of such activities? And do users respect the licenses of authors who tend to license their works more liberally?
- How can we engineer communities of digital media producers based on mutual trust and incentive structures leading to socially and economically desirable behavior?

Related work

A lot has been written on the use of open licenses and open standards for the production of software or digital content. A collection of related writings is provided in [3], while [4] provides several examples of collaborative, user-driven innovation, with an emphasis on software and physical goods. A more generalized treatise of peer-based production and its potential for the transformation of our culture and society is given in [5]. The power of remixing as a vehicle for creative expression is beautifully expressed in [6], although there are no quantitative analyses that we are aware of illustrating this power in practice.

For our study we use tools developed for social network analysis (SNA) [7], a field that is growing in importance and applications as the Web is becoming increasingly 'social' and 'participative'. For the visualizations presented herein we use the Netdraw application [8] developed for the visualization of social networks. A social network is a graph consisting of nodes typically representing individuals or organizations and edges representing ties between them. Ties are derived often from explicit relationships between individuals, as evident in patterns of (usually verbal) communication, i.e. emails, forum posts, telephone calls, etc. Several techniques have been developed for the study of such networks and while there are many traditional sociological and organizational behavior applications of SNA (e.g., to identify social cliques, professional networks within a firm, or between firms, etc.), SNA has also been successfully applied to the study of other types of 'networks', and of particular relevance to our study are what we may call *attribution networks*, as they are defined by some form of explicit attribution (e.g., scientific papers citing prior work,

[1] Recent Apple/EMI announcement of DRM-free iTunes Plus downloads and Amazon's new DRM-free MP3 store.

patents citing related inventions, etc.). A collection of key articles on SNA, with examples of several applications, including some related to 'attribution networks' is given in [9]. Another related application of SNA is the study of *affinity networks* [10] where implied ties between individuals are discovered through the mining of user data for similarities in people's activities or in their profiles.

Our study is also focusing on indirect relationships created through the citation of prior work rather than on direct communication. However, the links we study are not formed just by simple attribution, but by the actual incorporation of parts of somebody's creative work into a new work (through remixing, or, more generally, reuse). Our hypothesis is that such links create a unique type of tie between individuals, and perhaps create *stronger* ties, compared to simple attribution/citation, as the process of remixing of a creative work is much more involved than the process of making a citation in an academic paper, and furthermore, the actual works are vehicles of personal expression and thus, one may argue, more particular to the individual, more *personal*, compared again to an academic paper or a patent.

Although not related to social networks[2], content-wise, the only visualization that we are aware of that is closely related to the network diagrams we present herein, is the one produced by Jesse Kriss[3] to illustrate the process of sampling in the recording industry. The visualization shows a timeline of original recordings and a timeline of albums with samples. When the user clicks on a newer album with samples the links to all past recordings which were sampled for this album is presented, even at the level of individual tracks on the album. The visualizations we will present herein are static and not interactive, but provide an overview of all links created between earlier and newer works and across more than one generation of remixing. We are also able to produce visualizations at the level of authors and not only of songs or albums. We do not include time information yet, but we plan to do so in the future, so as to visualize and study not only the structure, but also the *dynamics* of creative reuse.

Empirical analysis

In an attempt to answer some of the questions we listed in the introduction, it is valuable to start from an empirical study of existing online communities. Based on the observations we make from the study of user behavior in such communities we can then proceed to formulate hypotheses regarding the motivations behind online sharing and remixing behavior and the users' subjective evaluations of the trade-offs involved in participation. Our ultimate goal is to produce a theory and methods (analytical and simulation-based) for analyzing the value of participation and creative reuse.

We started our exploration of creative reuse by collecting a rich dataset describing all uploaded material on the *ccMixter* online music community[4] which numbers 1,850 active users (actively engaged in content production and remixing – about 18% of the community's total registered users). The *ccHost* software running on the ccMixter site explicitly tracks the reuse of content by members, while also informing members and viewers about the terms of the Creative Commons licenses attached to all uploaded items.[5]

ccMixter started after Wired magazine published a CD with CC-licensed material from mainstream artists, inviting others to remix this material legally.[6] The founders of *ccMixter* wanted to create a community that would leverage this material but would also be a model for future communities, a sandbox of sorts where new ideas and tools enabling reuse can be tested. This com-

[2] For a collection of visualizations of networks of various types, including social networks, see http://www.visual complexity.com
[3] The History of Sampling, visualization by Jesse Kriss, available at http://jessekriss.com/projects/samplinghistory
[4] Community website: http://www.ccmixter.org
[5] Besides ccMixter, some user communities dealing mainly with music production and remixing are: JamGlue, SpliceMusic, Kompoz, YourSpins, and AcidPlanet.
[6] Wired Magazine, Issue 12.11, available online at http://www.wired.com/wired/archive/12.11/

munity therefore provides us with the unique opportunity to observe this “remix culture” enabled by the participative web with much more precision than we otherwise would. In the rest of this document we collect some of our early observations. Many of these observations have not been fully tested yet, but are discussed here for early communication and documentation purposes.

Structure and size of reuse network

The community had collectively produced an impressive 7,484 music items at the time of the data collection (June ’07). This is a very respectable output size if we take into consideration the moderate size of the community and the fact that the production of a music sample or complete piece (even if it is a remix) is generally more time consuming than the taking of an amateur photograph or the creation (or editing) of a Wikipedia entry. Interestingly, remixing accounts for more than half of the total production volume (3,982 items, or 53%), even if about 60% of all uploaded original music pieces (2,150 of 3,502) never get remixed[7]. This is already suggestive of the central role that reuse can play in digital media production.

Figure 2 on the next page shows the *content view* of the *reuse network* of the *ccMixter* community. We define a reuse network as a directed graph consisting of nodes representing entities (authors or content) connected by arcs representing reuse relationships. Figure 1illustrates this relationship.

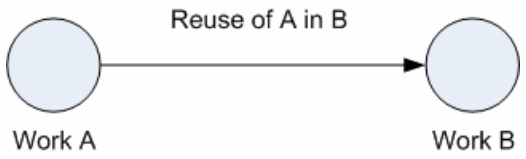

Figure 1: Reuse at the level of works

In the *content view* each node represents a single uploaded work (i.e. a complete music piece, an audio sample, or a sample pack) and arcs point from source works to derivative works. Node size in Figure 2 is relative to a node’s out-degree (i.e. number of remixes). Reuse is a many-to-many relationship, as a work may have several adaptations (we use the terms derivatives and adaptations interchangeably) and may itself be a derivative combining several prior works. Firstly we notice in Figure 2 that the network of reuse is very dense, which is indicative of the extent of reuse in the community. Also, reuse does not occur in isolated parts of the network; almost the entire graph is connected, except for the small nucleus in the center of the figure, which consists of a set of small independent reuse networks, as revealed by a components analysis. Hence, reuse apparently helps forge links that span the entire community, whereas initially we expected to see a set of smaller islands of reuse formed by users with similar taste or interests.

Nevertheless, as the sizes of the nodes in Figure 2 suggest, there exist regions of very high reuse within the network. These are typically generated by competitions organized by the community where recordings of well-known artists are (legally) uploaded to the website for a time-limited remixing competition. Such competitions clearly create additional incentives for reuse, though it appears that most of the content produced in this manner is a one-time effort and is not reused subsequently in any other works. We can postulate based on this initial observation (with a more detailed analysis still pending) that extrinsic incentives lead to higher reuse numbers but this reuse is short-lived and does not extend much beyond the confines of a competition. Hence we observe mini star topologies within the greater network.

[7] ccMixter also pools CC-licensed content from other communities, such as Freesound, and makes it available to its members for remixing. Our initial analysis in this document includes only content uploaded by the ccMixter community. We are working towards producing a more complete view of all content and will update this document accordingly.

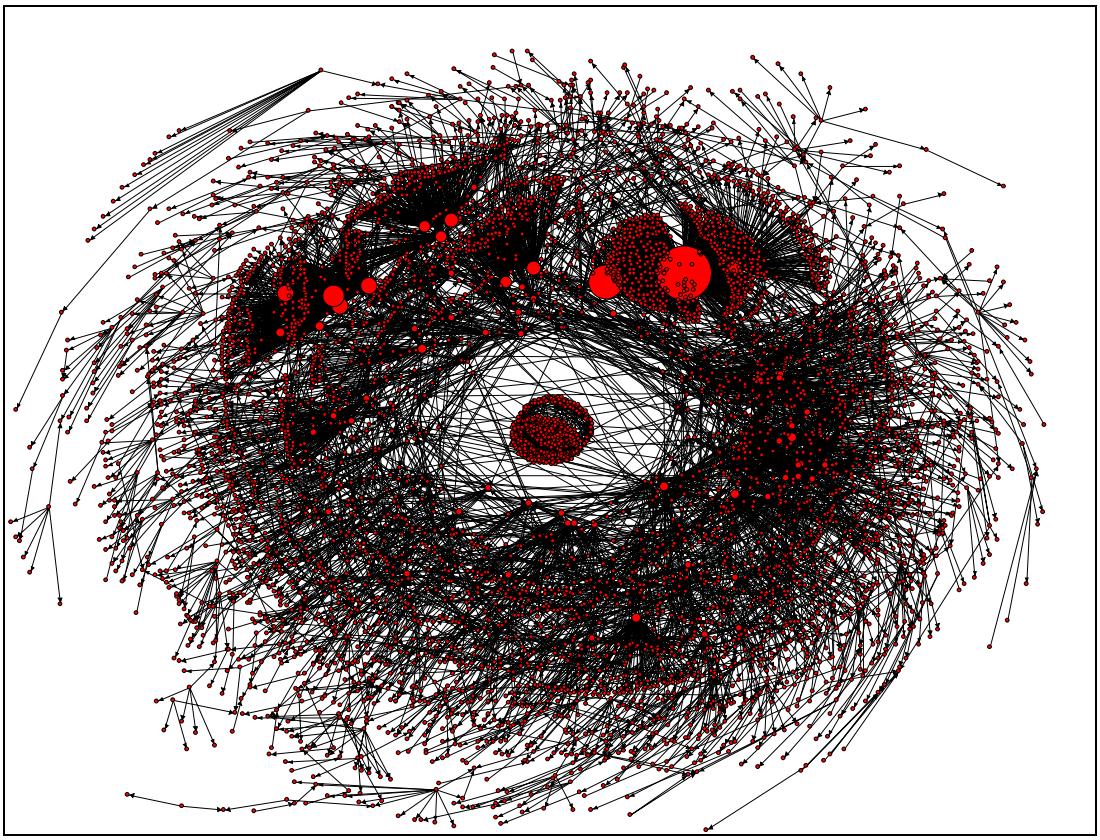

Figure 2: Reuse network – Content view (isolates removed)

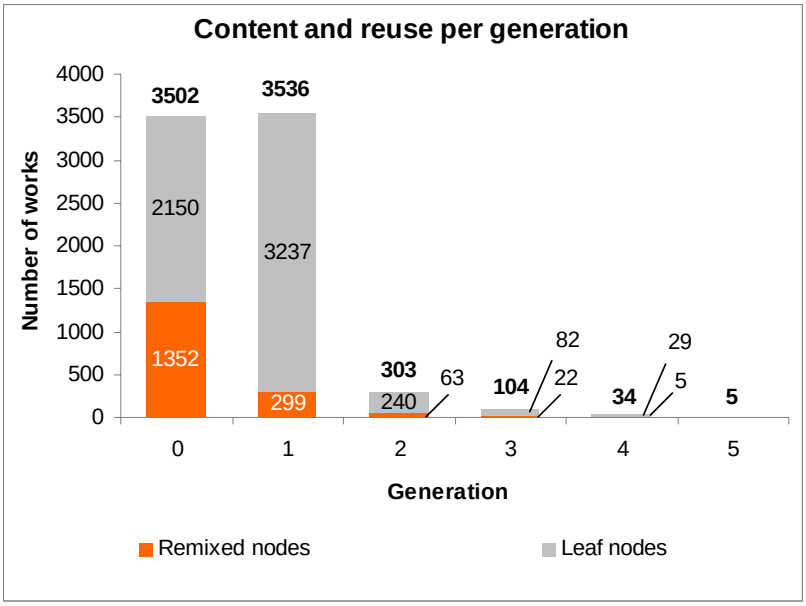

Figure 3: Number of works in the *ccMixter* community per generation of reuse

We are now in a position to ask: What is the *depth of reuse*, i.e. the maximum number of consecutive remixes in the network? Figure 3 shows the number of works per *generation of reuse*, with generation 0 representing original works. We notice that, surprisingly, the maximum depth of reuse in the network is 5, i.e. there are (5) works which are the products of 5 consecutive remix-

ing steps. Nevertheless, as we expected, the extent of reuse drops significantly in generations >1. Most users wish to remix original works, which is consistent with what we observe generally in the music industry (it is not common to produce remix albums of remixes). Also, we must take into account the fact that some Generation 0 content consists of individual samples/tracks, e.g., *a cappella* recordings, which lend themselves to reuse in multiple mixes. Content in subsequent generations is "mixed down" and therefore becomes harder to reuse. The relationship between *modularity* and *reusability* is also discussed in [11]. The striking number of works in generation 1 is largely due to the competitions, so we also plan to produce the same graph independently for competitions and for the network without competitions in a future version.

Figure 4 shows the cumulative percentage of works having been remixed X number of times, for each generation. First of all, this graph clearly shows that the relative frequency of reuse for a work follows an exponential distribution for each generation. Generation 0, i.e. originals, exhibit a very long tail of reuse – even if most originals (over 60%) are never reused, about 30% are re-used between 1 and 8 times, and it is not uncommon to find originals with much higher numbers of reuse (even outside competitions). This is less the case for subsequent generations of reuse, with a smooth decrease per generation. Generation 1 is unique again here in that it exhibits lower numbers of reuse compared to generations 2 and 3 (and is closer to generation 4 in Figure 4), because of the distorting effect of competitions, leading to over 90% of generation 1 works never being used again.

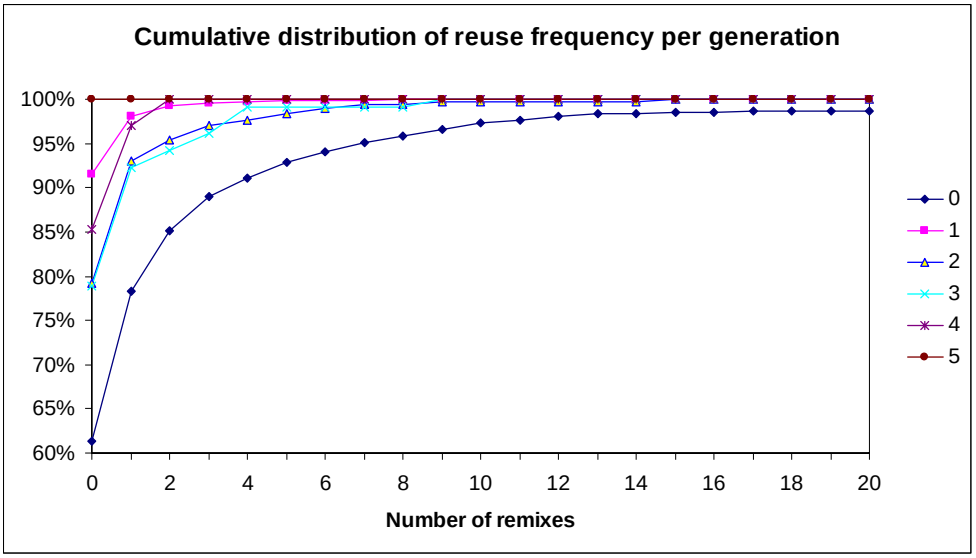

Figure 4: Cumulative distribution of reuse showing long tail

Another simple means of understanding the structure of the graph of Figure 2 is by examining the in- and our-degrees of the entire reuse network. These are shown overlaid in Figure 5 on the next page. We notice that, for lower degrees, in-degree is higher than out-degree, while the converse is true for higher degrees. This is because the reuse network consists mostly of 1-to-N relationships rather than N-to-1, or, in other words, it is occurs more frequently that one work becomes the subject of multiple remixes instead of a single remix utilizing multiple works as sources. We believe this will be a key characteristic of any reuse network, as it is generally more common and perhaps also easier to reuse one work in multiple contexts than it is to combine multiple sources into a new coherent work. It will be interesting to investigate whether this holds true for all media types or also for other types of reuse outside the domain of digital media, or, more broadly, cultural works.

Furthermore, if we examine the joint distribution of in- and out-degrees we will see that only works with low in-degree exhibit high out-degrees. In other words, the more source works a derivative is using, the less likely it is to be further reused, irrespective of whether it is a generation 1 or higher

work. This is shown in Figure 6. Combining this with our earlier observations, we can say that the attractiveness of a work, at least with respect to its attractiveness as source material for reuse, is decreasing in the *depth* and *breadth* of reuse of the work. The more "derivative" a work is, either because it is the product of many subsequent reuses, or because it is itself reusing many sources, the less likely it is that this work will be reused in future generations. Hence, we have shown that reuse may have the ability to greatly boost the output (and possibly the diversity of the output) of a community of authors, but is limited by a natural force that constrains both its depth and breadth.

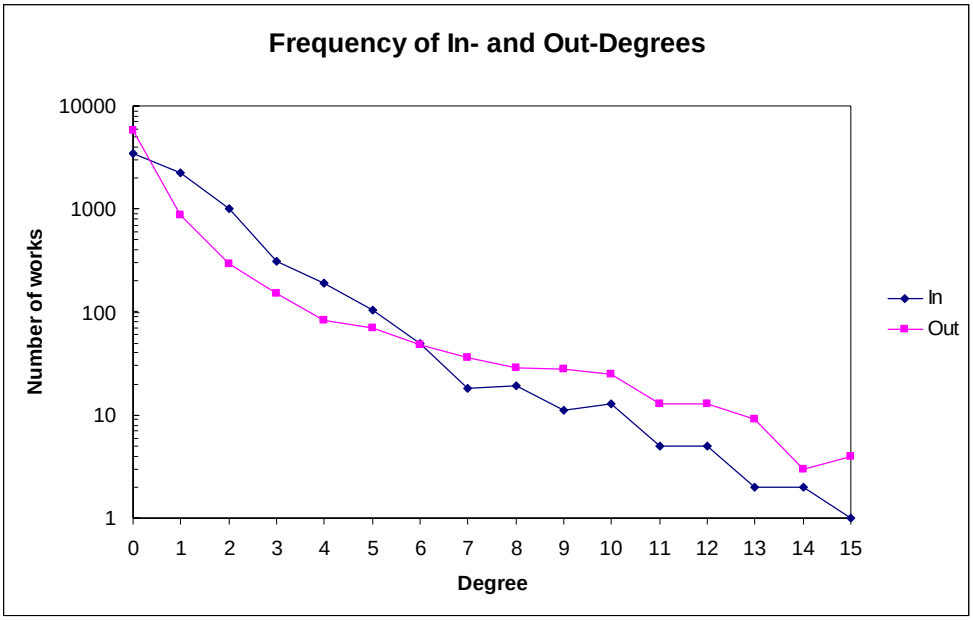

Figure 5: Frequency of in- and out-degrees in reuse network (log scale)

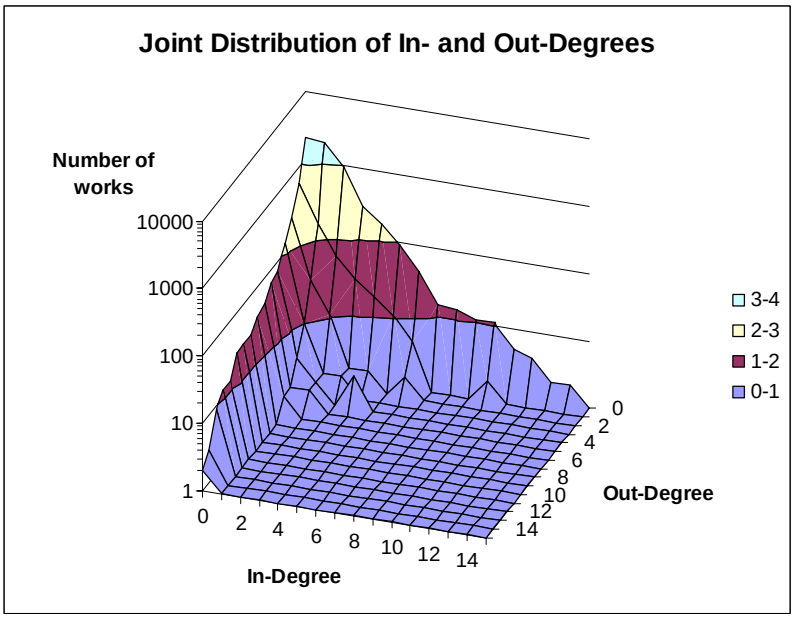

Figure 6: Joint distribution of in- and out-degrees (log scale)

Licensing and social enforcement

Another interesting aspect of reuse networks is that they provide us with the ability to observe how authors treat each other's works, and one dimension of this is licensing. As already men-

tioned, in the *ccMixter* community all works are licensed under Creative Commons licenses, some of which are more liberal, while others are more restrictive. When an original work is licensed under a particular license, derivatives of that work can only be legally licensed using the same, an equivalent, or a more restrictive license. Licensing under a more liberal license would violate some of the conditions of the original license. Also, when reusing multiple works, all of the licenses of these works have to be respected in the licensing of the derivative work, and this can sometimes be a source of confusion and complications as it is not always clear (or even possible) to respect the licenses of all source works.

Licensing of derivatives can thus be tedious and in general reuse can lead to more restrictive licensing. This narrowing of the licensing is another force limiting the depth and breadth of reuse, as more restrictive licensing will generally make subsequent reuse harder to accomplish from a legal standpoint. This narrowing may be voluntary on the part of the authors of the derivatives, where such an author may choose to be more protective of his/her work than the author of the original was, or may be involuntary, in cases where the reuse of multiple source works in one derivative work forces the derivative's author into more restrictive licensing.

Interestingly we have thus far observed that in ccMixter (a) authors of derivatives tend to respect the licenses of the works they reuse, and (b) in cases where they could legitimately license their derivatives under more restrictive terms, they generally do not. Upon closer inspection we found out that this is primarily the result of an ingenious licensing mechanism implemented by the site administrators. Every author of a remix must state the sources used in the derivative work. As the license of each source work is stored in a database, the website will automatically select an appropriate license for the remix. Thus license compliance is ensured.

According to the site founder and main operator, Victor Stone (who is also an active member of the community), this was implemented to improve the ease of use of the website, as users would sometimes be at a loss when having to choose an appropriate license. However, what is really interesting about this mechanism is that it is not based on a technical protection measure (DRM) but instead relies heavily on the users' correct and honest declaration of the sources of their work. Authors of remixes will naturally only use sources which they deem valuable for their work or which stem from other authors whose work they generally look up to. Outside the context of a community remix authors may not care to give full attribution (copying and reusing content without attribution is common on the Internet), but inside a community an author's status depends partly on the authors he/she is associated with. Remixing provides such an association and users will therefore be motivated to provide full attribution. We might call this an *incentive compatible social enforcement mechanism* because authors are intrinsically motivated to declare the true sources of their work and the community can utilize this information to ensure that author rights are respected.

This may well provide a new paradigm for dealing with author rights in a post-DRM era. Many efforts at introducing DRM for music have been unsuccessful and even in cases of relatively successful models, such as Apple's Fairplay solution, the industry has started moving away from DRM due to the costs it incurs and the inconvenience it causes for users. Incentive compatible social enforcement may be the answer. With the aid of a proper incentive structure users will be naturally motivated to respect authors' rights and this is central to the building of trust through reciprocity in a creative community. We will look into this in more detail, to determine exactly how this type of "social enforcement" may function in larger communities and possibly also outside the context of remixing.

Author network

Finally, it is interesting to group works by author and reproduce the graph of Figure 2 as a network of authors linked by the act of reuse. This type of relationship is depicted in Figure 7. The

result is shown in Figure 8.[8] The size of the nodes in Figure 8 is relative to the number of uploads (original contributions or remixes) that each user has made. We observe that most of the authors participating in competitions (clearly distinguishable again as nearly distinct star topologies) are otherwise not active members of the community. Competitions therefore have the ability to attract more members into a community and thus enhance its output, but these users do not necessarily become involved in any other activities besides one short-lived competition, so their contributions are likely not of much value to the community in the long term.

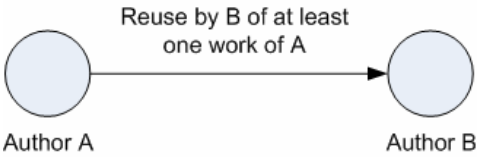

Figure 7: Reuse at the level of authors

Network density is highest in the center of the graph, where several authors are reusing each other's works, creating a complex mesh of adaptation and experimentation. This is in a sense the core of the community, and although more analysis is warranted on this point, we believe that Figure 8 strongly suggests that the creation of artificial and temporary incentives for participation is not of much value to the long term sustainability and well-being of the community (even though it does contribute significantly to the total output of the community).

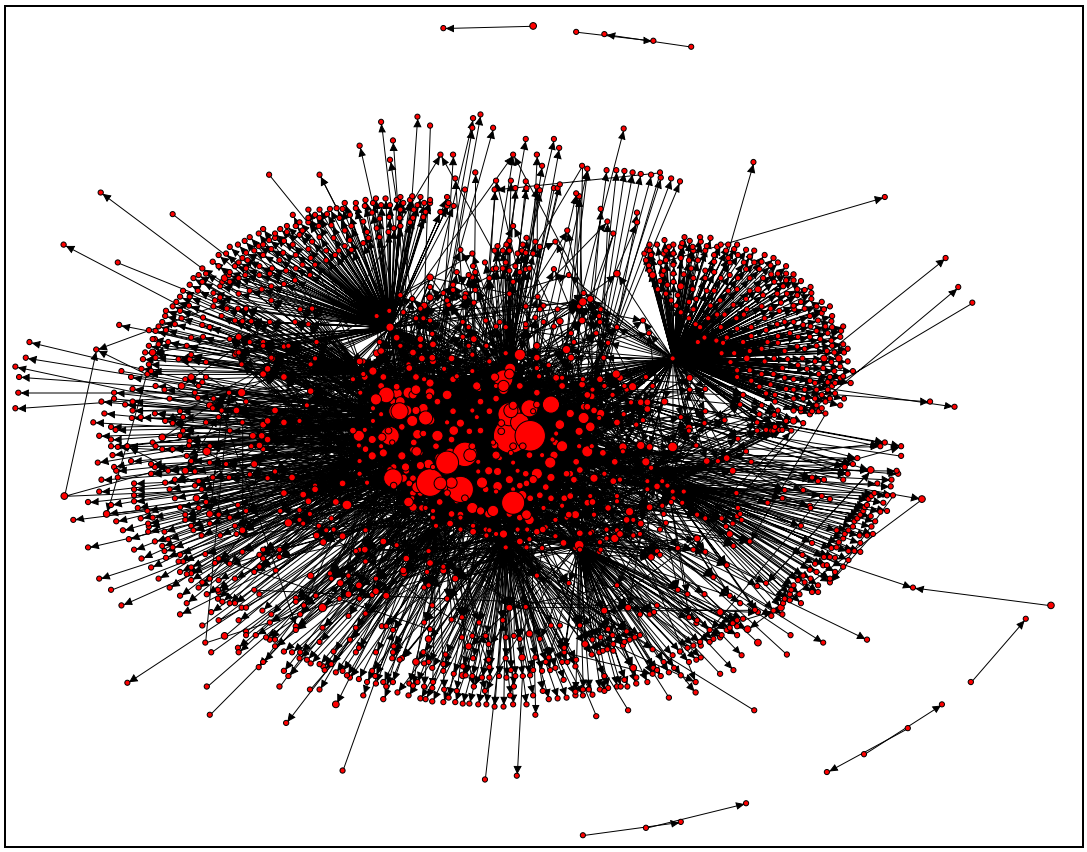

Figure 8: Reuse network: Author view (isolates removed)

[8] This Figure is based on a January 2008 dataset, whereas the data presented in previous parts of the paper is from a June 2007 dataset.

It will also be interesting to examine this as a social network consisting of *implied* relationships between authors, as opposed to networks created via *direct* communication (e.g., phone conversations, emails, internet chat), and also compare and contrast cultural reuse networks to other such implied relationships, such as those generated by paper citations in the academic literature. We believe that the repurposing of an author's creative output by another author creates a strong link between the two authors, as, through the creative process, they begin to share a common context that is very personal, even if they may not have engaged in any direct verbal or written communication. It will be therefore very interesting to assess the importance of creative reuse in improving cultural, cross-border understanding and also in improving a society's overall media literacy and generalized trust levels.

We are currently working towards a comparison of reuse relationships (indirect communication) and user forum contributions (direct communication) in the ccMixter community. Figure 10 shows the network of authors as defined by their communications in the user forum of ccMixter, with nodes sized relative to the number of posts per author. The semantics of the links in Figure 10 is shown in Figure 9. We notice that the structure of the network of Figure 10 is very different to the one in Figure 8, which implies that the relationships built through direct online communication are very different to the relationships built around content production.[9]

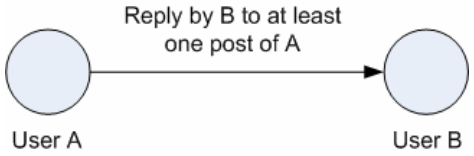

Figure 9: Communication links between users

As is clearly shown by Figure 10, one user is at the center of most forum interactions and has produced a significantly larger amount of posts compared to other users. Not surprisingly, this is the community's founder and main operator, Victor Stone. It is possible that the same user is also a key content contributor (of original uploads or of remixes), but before even going into deeper analysis, a cursory look at the data visualizations will suggest that the reuse network is depending on a larger set of key players that extends well beyond the founder. In subsequent versions of this document we plan to explore in more depth the interplay of online creative reuse and online communication in the formation of community relationships.

Summary and outlook

We introduced the concepts of a reuse network, reuse depth and reuse breadth, to organize and characterize relationships created through the process of remixing within a community. We further provided some of the first visualizations of our contemporary remix culture and initial thoughts and observations on two different views of a reuse network, the content view and the author view. Reuse can span multiple generations of works and has the potential to at least double the output of a community, although the propensity to reuse decreases rapidly with the breadth and depth of reuse, so there is a balancing force between original works and adaptations. Extrinsic incentives, as introduced through competitions, can boast the output of authors and attract more authors to a community, but may not be as valuable in the long run as most competition works are never used again and most new authors appear to limit their participation to the timeframe and scope of the competition.

We believe that the investigation of the paths that digital content follows online (which we might call 'cultural flows') and how the actions of authors and audience affect these paths will be a very exciting research area which will helps us develop a theory of online participation and reuse, as

[9] Figure based on January 2008 data.

well as tools for assessing the value of content and the value and "health" of a community. Wired magazine editor Chris Anderson writes in his influential book "The Long Tail" [12]:

> *"Why do they do it? Why does anyone create something of value without a business plan or even the prospect of a paycheck? The question is a key one in understanding the Long Tail"*
>
> And:
>
> *"Understanding the diverse incentives that can motivate the creators of such content becomes essential in finding and encouraging it."*

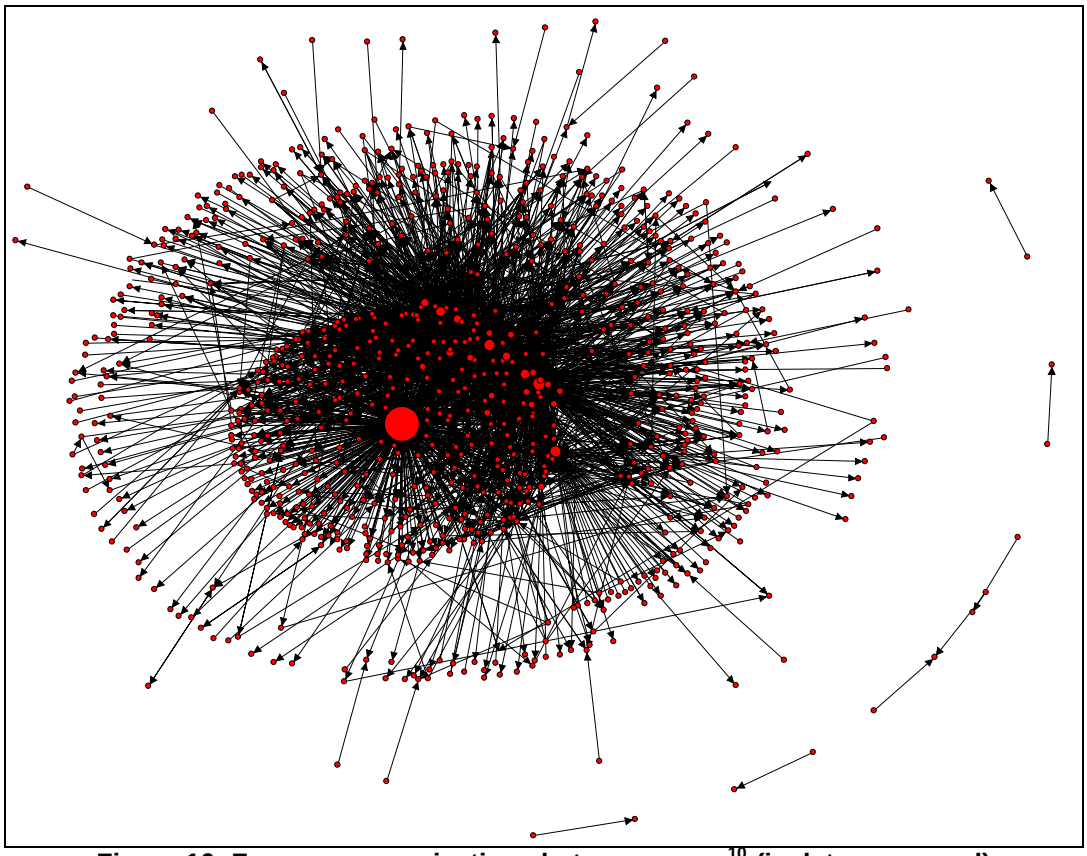

Figure 10: Forum communications between users[10] (isolates removed)

It will be useful to incorporate more elements of social network analysis, especially when investigating relationships between authors, to identify, among other things, which authors are key players in the community. Also, novel metrics will have to be developed to measure the unique properties of reuse networks, e.g., a measure of *diversity* will be useful in assessing the effect of reuse on not only the quantity, but also the quality and variety of content that is produced. We also plan to build a software tool for the simulation of online communities. This will help us study the dynamics of these communities from birth and in a controlled environment.

The simulation, visualization and analysis of remix culture must be placed in the broader context of what Professor Lev Manovich, author of "The Language of New Media", calls *cultural analytics*, i.e. the use of very large 'cultural' datasets, quantitative methods and new visualization technology to map and monitor the very creation and evolution of our modern digital culture. Besides

[10] Here 'users' refers to authors and non-authors (ccMixter users who may be participating in forum discussions without having uploaded or remixed a song).

serving as a valuable resource for researchers in the Humanities or as material for new media art installations, such visualizations will be utilized in identifying levels of activity and opportunities to invest in digital media production and distribution. This is very similar to the way that we use business analytics to assist us in business decision making, customer relationship management and operations management.

Indeed, besides the research and educational value that these maps of cultural flows will have, we can identify several business applications. For instance, organizations wishing to jumpstart user communities (for content creation, networking, technical support, or other purposes) face several issues with respect to the community setup, attracting the right members, creating incentives for participation, obtaining desirable output, etc. Also, the financial valuation of existing and successful communities is a challenging problem whose importance is increasing with the number of recent high profile acquisitions of social networking sites. The lack of appropriate models and tools for assessing and monitoring the value of a community and the content it produces is a serious shortcoming for any potential investor. As our media landscape is becoming increasingly dependent on the contributions of highly distributed and diverse teams of individuals for the creation of complex media products, it will become essential to produce a suite of media analytics tools, which will assist in the (eventually real-time) monitoring of the paths that digital content follows online.

References

[1] Cheliotis, G., Chik, W., Guglani, A. and Tayi, G. *Taking Stock of the Creative Commons Experiment*, 35th Research Conference on Communication, Information and Internet Policy (TPRC), Arlington VA, Sep 28-30, 2008.

[2] Lessig, L. *Free Culture: The Nature and Future of Creativity*, Penguin, 2004.

[3] Gosh, R.A. (ed.), *CODE: Collaborative Ownership and the Digital Economy*, MIT Press, 2006.

[4] Von Hippel, E., *Democratizing Innovation*, MIT Press, 2006.

[5] Benkler, Y. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, Yale University Press, 2006.

[6] Miller, P.D., *Rhythm Science*, MIT Press, 2004.

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

[8] Borgatti, S.P. *NetDraw: Graph Visualization Software*, Harvard: Analytic Technologies, 2002.

[9] Newman, M., Barabasi, A.-L., and Watts, D. (eds.), *The Structure and Dynamics of Networks*, Princeton Studies in Complexity, Princeton University Press, 2006.

[10] Smith, M., Giraud-Carrier, C. and Judkins, B. *Implicit Affinity Networks*, 17th Workshop on Information Technologies and Systems (WITS), Montreal, Dec 8-9, 2007.

[11] Manovich, L. *Remixing and Remixability*, available online at http://www.manovich.net on Jan 10, 2008.

[12] Anderson, C. *The Long Tail: Why the Future of Business is Selling Less of More*, Hyperion, 2006.