

An Analysis of the Social Structure of Remix Culture

Giorgos Cheliotis
Communications and New Media
National University of Singapore
Blk AS6, #03-41, 11 Computing Drive
Singapore 117416
gcheliotis@nus.edu.sg

Jude Yew
School of Information
University of Michigan
1075 Beal Ave.
Ann Arbor, MI 48109-2112
jyew@umich.edu

ABSTRACT

We present findings from our study of a music sharing and remixing community in an effort to quantify and understand the structural characteristics of commons-based peer production for products of aesthetic/cultural or entertainment value. We also provide a normative perspective on the strategies that such communities should employ with respect to the use of 'remixing contests', which are popular means of attracting new user-creators to the community and boosting its creative output. Until now research has shied away from the quantitative study of what lies at the heart of this 'remix culture', i.e. *remixing*, presumably because of the difficulties inherent in attaining relevant large datasets amenable to numerical analysis and an early focus of research efforts on communities whose products serve a more functional purpose (e.g., open source software), rather than aiming at entertainment or personal and artistic expression. This paper contributes to the literature of social network analysis of online communities, the literature on commons-based peer production, and the research agenda of cultural analytics.

Categories and Subject Descriptors

H.5.3. [Information Interfaces and Presentation (e.g., HCI)]: Group and Organizational Interfaces – collaborative computing, computer-supported cooperative work, theory and models, web based interaction. J.0 [Computer Applications] General.

General Terms

Measurement, Human Factors, Theory, Legal Aspects

Keywords

User-generated content, online community, creative reuse, social network analysis, remix culture, creative commons, incentives

1. INTRODUCTION

Much has been written about the birth of a new 'remix culture' on the Internet and how Web 2.0 technology and user-generated content will allow amateurs and even professionals to collaborate with each other across organizational or geographic boundaries [2,4,11,13,20]. Yochai Benkler has described this phenomenon as "commons-based peer production" (CBPP) [5]. Characterized as a means of organizing loosely connected individuals to openly share resources and cooperate without traditional hierarchy nor financial compensation, CBPP has been touted as a way of harnessing large numbers of highly distributed skills and human creativity towards a collective output or product. The usual suspects that highlight the success of CBPP are highly prolific examples such as

Wikipedia, Slashdot and Linux [5]. However, these are not the only examples of CBPP. What is missing, and what this paper attempts to cover, is a description and understanding of CBPP models that are focused on entertainment, creative reuse and cultural expression.

Such creative CBPP efforts often lack the production of a unified collaborative artifact or output. Rather, communities falling under this category of CBPP are motivated by the objectives of open sharing and reuse: to post original content and to create derivative works based on content made available by others. This ability to share and reuse content, while not focused on a common output, is bound together by some common objectives:

- Personal expression through the creation of content [4]
- Building social relationships through the creative process [14]
- Furthering the practices of communities that revolve around creating and personalizing content through remixes and mashups [13]

In this paper we study the cumulative process of cultural production in such a community. ccMixer is an online community that revolves around the open and legal sharing of music. All content on the community is legally uploaded, copyrighted and licensed under Creative Commons. Based on this open sharing the community members engage in the activity of reusing this open content to create derivative works - more commonly referred to as 'remixes'. This community is of interest because it not only allows us to observe the participants' behavior as they share and creatively reuse content, but also allows us to track the evolution of this content. While ccMixer is not the only online community that exhibits reuse of some sort, it allows us to quantitatively study what lies at the heart of our 'remix culture', i.e. *remixing*.

By analyzing ccMixer, we are able to uncover specific remixing patterns and confirm some of our assumptions about the structure of reuse behavior in an online community. Our study also provides an understanding of the dynamics of cultural production in general. Cultural production, like all other forms of production, "stands on the shoulders of giants" [5]; it builds upon the past, either directly, with the citation or creative appropriation of existing material, or indirectly, with the sharing of tools, know-how and ideas. Cultural production, much like the production of scientific knowledge and innovation, is cumulative [21].

This analysis of ccMixer will also highlight the role that contests play in the structure and dynamics of a community. ccMixer, on occasion, will invite well-known artists to contribute content to be competitively remixed, allowing community members to show off their creative and collaborative skills. Other online communities

that focus on remixing music, such as Kompoz, Jamglue, Splice and Acid Planet, also make use of contests to motivate participation and attract new users. Popular musicians also increasingly use such contests as a means of promoting new albums and engaging with their audience. While desirable in terms of generating excitement and attention on remixing activity, we believe that the use of contests to motivate participation has broad ramifications for online communities in the long run. Contests introduce additional extrinsic incentives for participation and thus alter the incentive structure of the community, and consequently, its output. The results of our analysis of remix contests in ccMixer will be of great relevance to managers, administrators and designers of online spaces that seek to encourage participation, sharing and creative reuse. We chose to focus on music communities initially, as there is a longer history of creative reuse of recorded content in music and it is generally a more established practice for this medium.

This study is also situated within a broader research framework that has ties to the agenda of cultural analytics [14]. The ultimate goal of our investigations is to produce practice-informed theories and methodologies that allow us to analyze, predict and effectively intervene in the process of developing media in a more open and collaborative fashion.

In summary, this paper presents a social network analysis of ccMixer, to uncover the following:

- i. The structure and characteristics of ccMixer as a network of authors connected by the content they reuse between each other.
- ii. The structure and dynamics of reuse at the level of the content, where our emphasis is on works, not authors, and cultural production is viewed as a process of adaptation and remixing of creative works
- iii. The role that contests play in increasing user participation and community productivity through the introduction of extrinsic incentives to participate and contribute content.

2. RELATED WORK

There is a growing body of literature 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 [10], while [21] 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 [16], although there are no quantitative analyses that we are aware of illustrating this power in practice. Remixing is also the subject of a recent book that has received significant media attention but is focusing more on the battles surrounding the legality of remixing under copyright law [13].

For our study we use tools developed for social network analysis (SNA) [3,6,22,23,24], a field that is growing in importance and applications as the Web is becoming increasingly ‘social’ and participative. A social network is a graph consisting of vertices typically representing individuals or organizations and edges representing ties between them. Many other applications of SNA-related techniques exist, but we will be focusing mostly on social

ties. Here, however, we focus on the ties created through the reuse of media content, i.e. ties that may be occasionally supplanted by verbal online/offline communication but that are formed mostly on the basis of CBPP of digital media content.

Given that, to the best of our knowledge, this is the first study of its type, we will build on some seminal papers in the field to compare and contrast our findings with studies in related domains. A collection of key articles on SNA, with examples of several applications, including citation networks, which are similar to remix networks, is given in [19]. Other related applications of SNA are found in the study of expertise and knowledge sharing networks [1,24] where implied ties between individuals are discovered through the mining of user data for similarities in people’s activities or in their profiles.

With respect to citation networks, we should note that the ties we study in this paper are not formed just by simple attribution, but by the actual incorporation of parts of someone’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. The process of remixing a creative work is much more involved than the process of making a citation in a book or academic paper. Furthermore, remixes are vehicles of personal expression. One may thus argue that these remixes are more particular to the individual, more personal, compared again to an academic paper or, say, a patent.

Although not directly related to the social networks literature, a relevant visualization of remix culture is the one produced by Jesse Kriss¹ to illustrate the history of ‘sampling’ in the recording industry. However, this study too, like most works on the subject of remixing, does not include any quantitative analysis of the remixing process.

3. NETWORK DESCRIPTION

The ccMixer community is known for being among the first to use Creative Commons (CC) licenses for the promotion of user-generated remixing activity. This community was also the first to use the open-source ccHost software, developed to explicitly track the reuse of content by members, while also informing members and viewers about the terms of the CC licenses attached to all uploaded items. The community was jumpstarted shortly after Wired magazine published a CD with CC-licensed material from popular artists, inviting others to remix this material legally. The founders of ccMixer wanted to create a community that would leverage this material but would also be a model for other communities, a sandbox of sorts where new ideas and tools enabling reuse can be tested.

Remixing in the community, like in many similar communities, occurs in two fashions: either *ad-hoc*, i.e. depending entirely on a user’s intrinsic motivation to remix somebody’s work, or organized through *contests*, which provide additional incentives for users to remix works within a given timeline. The fact that contests provide additional, extrinsic incentives for the remixing of works is of special interest as this is a very common method of

¹ The History of Sampling, visualization by Jesse Kriss, available at <http://jessekriss.com/projects/samplinghistory>

attracting new members to an online community and we wish to investigate their effect on the community.

Our analysis of creative reuse begins with a characterization of the activity in the online music community which numbers 2,145 active users (actively engaged in content production and remixing), who produced a total of 9,300 submissions. Active users, whom we will often refer to as ‘authors’, constitute 17% of the community’s 12,776 registered members. Of these authors, 1,698 (i.e. 79% of the 2,145) have remixed at least one submission of another author or have had at least one submission remixed by another author. The rest of the authors have uploaded music tracks which have not been remixed. The relatively high number of authors engaged in remixing is likely attributable to the nature and mission of the community, which is centered on the promotion of legal and free remixing through the use of Creative Commons licenses for all submitted content. However, in spite of the relatively high number of authors engaged in remixing, we will see that many authors also produce a lot of content which does not get reused by others.

In this paper we will focus our attention on the structure of the relationships created between works and between their authors through the reuse of creative content. Every time an author chooses to remix the work of another author, this creates a link between the original and the derivative work, as well as between their respective authors. We define a *reuse network* as a directed graph consisting of vertices representing entities (authors or content) connected by edges representing reuse relationships. Figure 1 below illustrates the convention that we will use for the representation of this relationship in the *content view* of the reuse network.

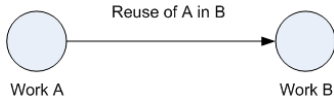


Figure 1: Reuse at the level of content

In the *content view* each node represents a single uploaded work (i.e. a complete music piece, an audio sample or a special effect) and edges point from source works to derivative works. Reuse is a many-to-many relationship, as a work may have several adaptations (we will use the common terms ‘derivative’ and ‘adaptation’ interchangeably) and may itself be a derivative combining several prior works. Table 1 provides an overview of the content network of ccMixer, organized according to *generations* of reuse. The generation a music piece belongs to is determined by the length of the longest path between the set of original works and that piece. So, an original work will belong to generation 0, while a work that is reusing two works, one of generation 0, and one of generation 1, will belong to generation 2. The rationale for using the longest path is based on the observation that the aforementioned hypothetical work would not have existed in its current form had it not been for both a generation 0 and a generation 1 source work: it should therefore be classified as generation 2. The classification of works in this manner allows us to examine the *depth* of reuse in such a network, which will complement our investigation of the *breadth* of reuse, the latter being determined by the distribution of out-degrees.

We can see in Table 1 that reuse depth in the ccMixer network is 6, which is surprisingly deep given that in popular musical culture

we do not often encounter remixes of remixes. Nevertheless, as we expected, the extent of reuse drops significantly in generations greater than 1. Most users wish to remix original works, which is consistent with what we observe generally in the music industry and in popular culture – it is easier to create derivative works from un-remixed material. Also, we must take into account the fact that some generation 0 content consists of individual samples or 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 [15] with similar conclusions.

We should note that the striking number of works in generation 1 is partly attributable to contests which were organized by the community at different points in time. Contest-related products consist of contest *entries* and contest *byproducts*. Entries are the official submissions by authors during the duration of a contest and in accordance with the contest’s rules, to be considered for the contest prize. For simplicity we also include under ‘entries’ the contest ‘source files’, i.e. the music pieces that users are called upon to remix within the context of the contests. Contest *byproducts* on the other hand are works which reuse contest entries (i.e. the source files or user-submitted entries), often mixing these together with other material from the community, but are not submitted as contest entries themselves. From the time that an individual music piece is uploaded to the community website as part of a contest, it is also made available to everyone who may wish to remix it outside the context of the contest and even after the contest has ended.

Table 1: Description of ccMixer content network

Network summary	Originals	Remixes by generation						Total
		1	2	3	4	5	6	
Nodes	4253	4501	381	113	43	8	1	9300
Original/Remix	4253	5047						9300
(% of total)	46	54						100
Contest nodes	57	1504	115	46	25	8	1	1756
(% of network)	1	33	30	41	58	100	100	19
Entries	57	1183	9	1	0	0	0	1250
Byproducts	n/a	321	106	45	25	8	1	506
Non-contest	4196	2997	266	67	18	0	0	7544
Original/Remix	4196	3348						7544
(% of total)	56	44						100

We notice in Table 1 that 33% of generation 1 remixes are contest nodes, of which the majority are actual contest entries, whereas the source material for all past contests on ccMixer was just 57 songs. Contests clearly then generate a great amount of remixing activity, in other words exhibit an extremely high reuse breadth. Perhaps surprisingly, they also contribute to the depth of reuse in the network, but only indirectly, through the community’s ad-hoc generation of contest byproducts, outside the context of the contests themselves. This happens because contest source files tend to come from well-known artists whose works enjoy continued popularity even after the expiration of the contest.

Overall, remixing accounts for more than half of the total production volume (5,047 items, or 54% of all works), even if

about 60% of all uploaded original music pieces never get remixed, as we will show later. This is already suggestive of the central role that reuse plays in the ccMixer community and is also indicative of the potential that reuse has for growing the output of an online community, although clearly the organization of contests also helps in boosting output volume.

Table 2 provides a different view on reuse per generation, by showing how many works are reused in each generation, leading to the next. We notice that 60% of all original works are never reused. This does not necessarily mean that they are of no value to the community, as they may be popular in their original form but not deemed very suitable as material for remixing. However, given the strong emphasis of the community on remixing, it would be unlikely for original works to be popular and *not* get remixed. This 60% therefore must include a lot of ‘waste’, from the perspective of the community, even if its production may have been worthwhile for the individual author. We also notice that only 9% of generation 1 works are remixed, much lower than the 40% of remixed works for generation 0, but also lower than the corresponding percentages of remixed works in subsequent generations. This is again attributable to contests which generate many generation 1 remixes and disproportionately fewer remixes in generations greater than 1.

It would thus far seem that contests tend to have a positive effect on the community only in the short term, in terms of providing additional temporary incentives for the production of more content, while having a negligible, if any, effect in the long term. They may even have a negative effect in the long term, because they detract from the normal, organic, growth of the community. We also do not know yet what their effect is on membership growth and what type of members tend to be attracted to the community through contests. We will provide answers to the above in the following sections.

Table 2: Overview of reuse activity by generation

Summary of reuse activity	Originals	Remixes by generation						Total
		1	2	3	4	5	6	
Reused	1691	392	71	24	6	1	0	2185
(% of total)	40	9	19	21	14	13	0	23
Not reused	2562	4109	310	89	37	7	1	7115
(% of total)	60	91	81	79	86	88	100	77
Total works	4253	4501	381	113	43	8	1	9300

Another means of understanding the structure of the content network is by examining the in- and out-degrees of nodes in the network. We provide the distributions of in- and out-degrees in Figure 2. Power laws [2,9] of the form $y=x^{-\alpha}$ provide very good fits for both distributions, for values of the parameter α close to those that have been used to fit observed distributions in many complex social and natural networks [9]. More tests are required to establish with certainty whether the degree distributions in ccMixer follow a power law [9], but the apparent long tail in the distributions is suggestive of the complex nature of a reuse network and may be attributable to preferential attachment [18]. In the context of remixing, preferential attachment would mean that works exhibiting a high degree of reuse become more attractive for further reuse. This concurs with observations outside online communities: many producers of modern music genres rely

heavily on the sampling of specific works to achieve a sound that is considered representative of the genre. The more frequently a specific sample is used, the more likely it is that more producers will wish to use the same recognizable sample in their works. This is a matter worthy of further investigation as it may prove valuable in understanding the forces that drive cultural production in general, but it is outside the scope of this paper, as our focus here is on the impact of contests on the output of an online community.

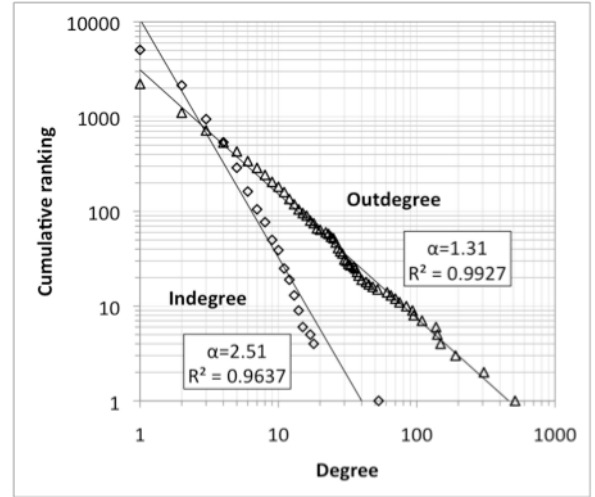


Figure 2: Cumulative distributions of content network in- and out-degrees

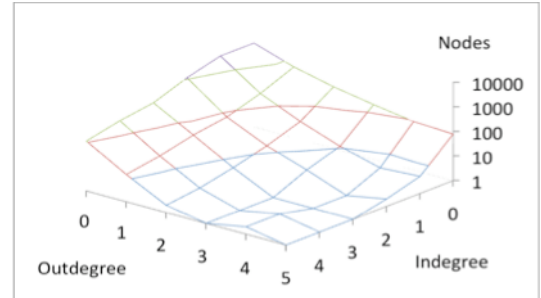


Figure 3: Joint distribution of in- and out-degrees (only degrees ≤ 5 shown, for better visual clarity)

We further notice with respect to Figure 2, 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 more frequent 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 a single work in multiple contexts, than it is to combine multiple sources into a new coherent work. 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 3.

Combining this with our earlier observations, we can conclude that the attractiveness of a work 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 natural, or perhaps ‘cultural’, constraints.

Finally, it is also interesting to group works by their respective authors, in order to produce the *author view* of the reuse network, i.e. produce a network of authors linked by the act of reuse. This type of relationship is depicted in Figure 4. This author network is a special type of social network, consisting of implied relationships between authors based on non-verbal communication through the reuse of each other’s musical ideas, as opposed to networks created via direct, verbal communication (e.g., phone conversations, emails, internet chat).

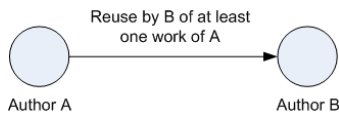


Figure 4: Reuse at the level of the authors

We believe that the repurposing of an author’s creative output by another author creates a potentially strong link between the two authors, as, through the creative process, they begin to share a common context that is very personal (as is the creative process itself), even if they may not have engaged in any direct verbal or written communication. We are still in the early stages of trying to understand this relationship, but there is strong evidence of a wider emerging recognition of the power of such ties. For example, it is becoming increasingly common for established musicians to offer one or more of their music tracks for remixing by their fans, in an effort to engage their audience in the creative process and through that, cultivate their relationship to the audience, as well as attract more fans who will appreciate the ability to participate in this manner.

The author network is smaller than the content network (2,145 instead of 9,300 nodes), as nodes representing works are grouped per author, but it is more complex, because in the author network it is possible to have loops and reciprocal relationships between authors. We will explain these differences in the following two sections, where we will analyze the structure of the author network and, where appropriate, compare and contrast it to the structure of the content network. With a combined analysis of these two complementary views of a reuse network we will be able to uncover the types of relationships formed through creative reuse, as well as the impact of contests on the social structure of these relationships.

4. COMPONENT (‘BOW TIE’) ANALYSIS

One of the most fundamental questions with respect to the members of any social network is whether they tend to occupy

different positions in the structure of the network, which may be representative of a difference in status, role, or influence. After careful examination of high resolution visualizations of both the content and author networks, we noticed the formation of a core of authors who seemed to keep the community together, while contests appeared to attract many peripheral authors who would not make any contributions outside the context of the contests. This motivated us to employ a bow tie structure analysis in order to identify the core and other peripheral components – a method that is most known for its use in characterizing the large-scale structure of the Internet as a bow tie shaped network [7].

This method of analysis categorizes nodes in a network according to one of the following components: i) a strongly connected core (SCC), ii) an in-bound link component (IN), iii) an out-bound link component (OUT), iv) tendrils and v) tubes. The core consists of authors who are highly interconnected with each other based on the sharing and remixing of each other’s work. The authors are strongly connected in this component of the bow tie structure because every author is connected to every other author in the core through the remixing relationships. The IN component consists of authors who upload content that is used by nodes in the core, but they do not use content from authors in the core. The OUT component consists of authors who are remixing works of authors in the core, but whose works are not used by authors in the core. Authors classified in tendrils connect to authors either in the IN or OUT components of the network but not to those in the core. These are authors who only remix the works of those in the IN cluster or who are remixed by authors in the OUT cluster.

The remaining authors in the network are categorized in the tubes component of the bow tie structure – tubes are nodes that create paths between the IN and OUT clusters while not being members of the core themselves. Finally, a certain percentage of nodes are disconnected from the rest of the network. These are authors who form small isolated ‘islands’, by virtue of being connected to one or more other authors, but not to the majority of authors in the network. In our analysis we will not pay as much attention to tubes or islands, as they form very small parts of the author network. See Table 3 for the results of the analysis.

Figure 5-A reproduces the results from the seminal paper by Broder et al. [7]. In that paper the authors express their surprise that a relatively small core, consisting of only 28% of all Web pages, was responsible for connecting all other pages in the network. Such pages act as hubs in the Web structure and play a key role in holding the Web together, as without them, half the Web would not be reachable from the other half. Although a direct comparison of these results to ccMixer would not be meaningful, we can make similar observations in the author network of ccMixer. First of all, the Core, IN and OUT components of the ccMixer network are all relatively small (see Figure 5-B). Only about 12% of the authors in ccMixer are actively sharing and remixing content with each other in the core component, while only 17% of the authors in the network are in the IN cluster and 20% in the OUT cluster. The aforementioned components appear relatively small because the author network has very large tendrils (see Figure 5-B and Table 3): 50% (or 841 nodes) of the authors are connected to either the IN or OUT clusters of the ccMixer network.

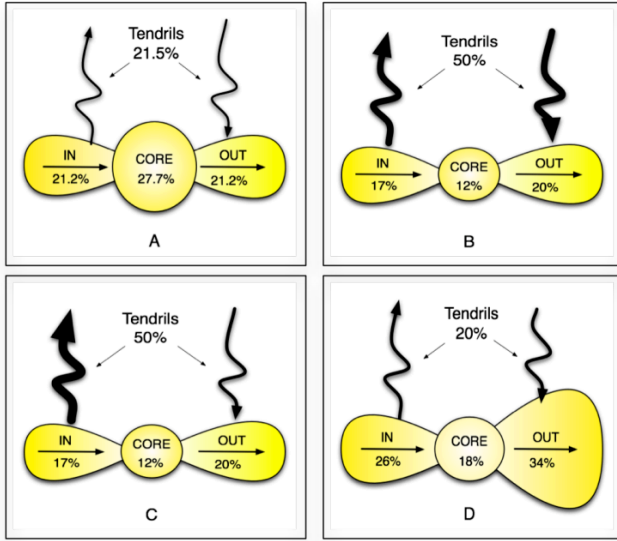


Figure 5: Results of bow tie analysis for (A) the Web, (B, C) the ccMixer author network and (D) the ccMixer author network with contest nodes removed

Table 3: Comparison of bow tie structures

Bow tie	SCC	IN	OUT	Tendrils	Tubes	Islands
ccMixer	12% (205)	17% (290)	20% (338)	49.6% (841)	0.4% (7)	1% (17)
No contests	17.8% (174)	26% (254)	34.2% (336)	20.1% (197)	0.2% (2)	1.7% (17)

One possible explanation for the oversized tendrils in ccMixer is based on the effects of contests mentioned earlier in this paper. Contest nodes consist of individual tracks or ‘sample packs’, typically from popular artists who donate these tracks to the contest so that they may be used by a large number of community members who otherwise do not necessarily remix content from authors in the core. In Figure 6 we show a typical link pattern for a prominent contest – in this case a contest to remix 2 tracks uploaded by a popular musician, Fort Minor. What the figure illustrates is that original contest nodes (i.e. the ‘source files’ uploaded to the community for remixing) have de facto zero in-degree – the contributing artist rarely becomes an active member of the community. Rather, these contest nodes have a large outdegree and, more interestingly, are connected to a large number of one-time remixers (the size of the nodes in the figure is in proportion to the number of uploads). Additionally, these ‘source’ contest nodes are also connected to some members of the highly interconnected core of the network (seen on the lower right hand side of Figure 6), which are also exhibiting a generally higher number of uploads (we will return to this point later).

Based on these observed patterns of connections for contests we are led to assume that contest ‘source’ nodes can only (or primarily) be allocated to the IN component of the ccMixer bow tie structure and that the large number of authors found in tendrils are one-time participants of contests. This would mean that the oversized tendrils would primarily lie on the side of the IN component. See Figure 5-C for an illustration.

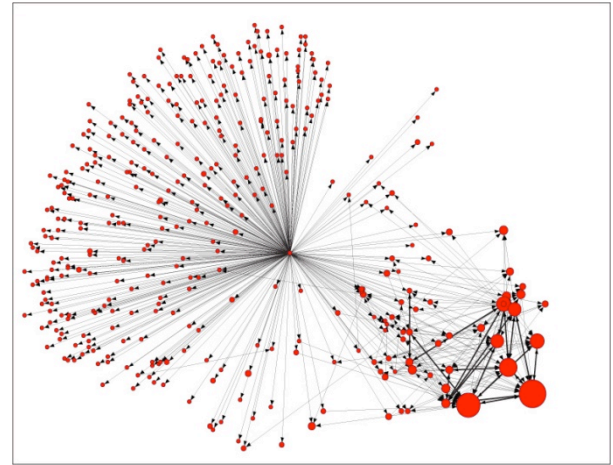


Figure 6: Ego network of Fort Minor contest

In order to test this assumption, we produced a view of what the author network would look like without contests. This was achieved by means of the following process: we removed from the content network all nodes deriving from contests, i.e. all contest entries, including the source files, and all contest byproducts (see definitions in previous section). Then we recreated the author network based on only the nodes that remained in the content network. The result of this process was that out of the 1698 author nodes in the ccMixer network, 718 authors were only associated with contests (directly or indirectly, i.e. from contest entries or contest byproducts) and were thus removed. We performed a bow tie analysis on the resulting network and the results can be seen in Figure 5-D and the last row of Table 3.

The results show that, indeed, contests did account for the bulk of the nodes in the IN-tendril cluster of the original ccMixer network. In general, removing contests caused the number of Tendrils to be reduced from 841 to 197 nodes. Also, the absolute numbers of nodes in the IN and Core clusters decreased somewhat (see Table 3). This validates the intuition put forward in Figure 5-C, i.e. that the large numbers of nodes in tendrils are a result of one-time remixers being connected to the contest nodes in the IN component.

What this highlights is that contests attract many new users and generate additional activity for existing members, but this participation tends to be one-time contributions and the majority of individuals who participate in contests tend not to participate further and are not connected with the core (and thus not connected with the majority) of the ccMixer community. These results serve to strengthen our suspicions regarding the value of such contests for the long-term sustainability of a community based on reuse. Another effect of removing contest nodes and conducting a bow tie analysis is that the relative percentages of the other clusters in the network have increased (see again Figure 5-D and Table 3). In fact, the Core and IN components now account for, respectively, 17.8% and 26% of the network structure, despite the fact that the numbers of nodes in these clusters have decreased somewhat. Given the reduction in tendrils and the IN component, the OUT component of the ccMixer network sans contests accounts for 34% of the network. The size asymmetry between the OUT and IN components, where the

former is larger than the latter, is suggestive of the asymmetry inherent in creative reuse, i.e. the fact that remixing is less ‘costly’, in terms of time and effort, compared to the production of original material that will be popular itself as a source for remixing (because remixes will reside more often in the OUT component, whereas popular originals generally will not).

The relatively small core in the ccMixer network is of value to the community because it helps connect about 80% of the community (in the contest-free network). Authors belonging to the core are indeed key drivers of community activity, in terms of remixes and in terms of the total amount of content they upload to the community, as can be clearly seen from Table 4. They also exhibit stronger ties and between them they form almost all reciprocal ties in the author network of ccMixer. Given that this component is relatively small, especially in the network as it stands with contests (it is only 12% of the total), communities may wish to question the wisdom of a strategy that is skewed towards growing the community’s size through high-profile contests and introduce strategies for the preservation and growth of the strongly connected core. Table 4 also displays the relative standard deviation of the metrics for the core and for the whole network, which is much lower in the core for degree and uploads. Two values are given for degree, corresponding to the relative standard deviations of in- and out-degree values (the average value is the same). The lower dispersion of values in the core is most likely attributable to the fact that authors in the core view each other as equal peers, thus also forming reciprocal ties.

Table 4: Comparison of core to network

SCC comparison	Core		Entire network	
	Value	Rel. σ	Value	Rel. σ
Reciprocal ties	114	n/a	115	n/a
Average tie strength	2.15	1.44	1.95	1.20
Average degree (in,out)	6.8	1.5, 1.3	2.9	2.8, 4.7
Average uploads	21.4	1.3	5.0	2.4

5. MOTIF ANALYSIS

The analysis thus far serves to highlight different structural elements of remix culture at the macro level and the effect of contests on that structure. It does not yet tell us much about the linking patterns between individual authors. We provided one example, the Fort Minor contest, but it would be more instructive to study the totality of the patterns of interactions that emerge in a remix network and how these are affected by the extrinsic incentives introduced by contests. To uncover these patterns of remixing we conducted a motif analysis of the author and content networks in ccMixer [17,23]. Specifically we focused on an analysis of the frequency of occurrence of all possible directed interactions (i.e. remixing relationships) between any 3 songs/samples in the content network and between any 3 authors in the author network. Motif analysis is premised on a comparison of the relative frequencies of occurrence of different linking patterns in a network to the respective frequencies in random networks with similar structural characteristics [17]. The differences in these frequencies we call ‘network motifs’. We can thus also ascribe a level of significance to the observed motifs in the network under study. Figure 7 lists the patterns that we will examine in this section due to their significance in the ccMixer

network. We will use this figure as a reference throughout this section (Feed B/F stands for ‘Feed Back/Forward’ as these patterns contain both feedback and feed forward loops).

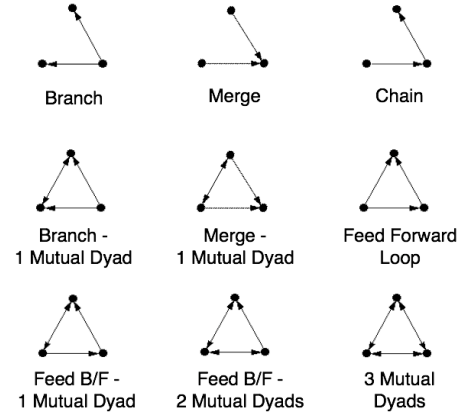


Figure 7: Patterns in remix networks

Table 5: Content network motifs

Motifs	Content Network		No Contests	
	Frequency	Z-score	Frequency	Z-score
Branch ^a	95.52% ^b	8.9	76.87% ^b	28.7
Merge ^a	3.83%	-8.9	19.90%	-29.2
Chain ^a	0.62%	-8.9	3.06%	-28.7
FF Loop	0.03% ^b	8.9	0.16% ^b	29.1

^a Most frequently occurring patterns

^b Significant motifs ($p < 0.05$) for patterns with frequency $\geq 0.01\%$

In Table 5 we show the 3 most frequently occurring patterns in the content network as well as those patterns that exhibit frequencies greater than 0.01% that are significantly different to the frequencies of occurrence of these patterns in similar random networks ($p < 0.05$). In this manner we can observe which patterns occur more frequently in a remix network as well as which patterns are perhaps not so frequent but nevertheless statistically significant. The intuition behind this distinction is the following: some patterns will naturally emerge more frequently in a remix network and this helps us characterize the network. However there are also patterns that are not as frequent, but which authors in a remix network exhibit a particular preference for. The former are important because they largely define the structure of remix networks (e.g., the in- and out-degrees of nodes) and the latter because they occur more frequently than one would expect within that structure.

We notice that the Branch pattern is dominant (in the sense of being the most frequently occurring pattern in the content network *and* significantly more frequent in that network than in a similar random network), with the Merge pattern a distant second. This is in agreement with our earlier discussions on the asymmetric nature of remixing and the plot of degree distributions in Figure 2. Consider for example James Brown’s ‘Funky Drummer’, which has the reputation of being the most sampled song ever. This track has been sampled by literally hundreds of other songs. If we were to plot these reuse relationships in the form of a network graph, we would obtain a very large number of Branch patterns.

The Chain pattern is the third most frequently occurring pattern, but with a much lower frequency than either Branch or Merge, which again agrees with our earlier observations regarding the length of reuse chains and the rapidly decreasing number of remixes per generation. Finally, Feed Forward loops will occur rather rarely, i.e. it is uncommon for the derivative of a song and the original song to be reused in the same music track, but given that this pattern is statistically significant (i.e. it exhibits a significant network motif according to the nomenclature in [17]), we can assert that this practice, even if rare, is nevertheless a (second order) characteristic of remix networks. There are no feedback loops or mutual dyads in the content network (the graph is acyclic) because remixing is strictly one-directional, i.e. it is not possible for a work to reuse its offspring.

When removing contest nodes, we notice that the frequency of the Branch pattern is reduced significantly, while the Merge pattern exhibits the largest gains. We can thus infer that in an ad-hoc remixing network, one that does not rely on contests, there is a lot more synthesizing of prior content into new derivatives that reuse more than one piece of source material. The relative emphasis that a community places on contests will thus clearly have implications for the type of content that the community will produce. In the absence of contests most new tracks will be remixes of a single source, and in that sense, close substitutes for one another, but many authors will mix different sources together, which leads to greater cross-pollination between works, authors genres and musical ideas. On the other hand, a community intent on growing through contests will create a lot more works that are close substitutes for one another and interest in these derivative works will likely decline quickly as the number of substitutes becomes very large.

Table 6 displays the most important patterns found in the author network, where MD stands for ‘mutual dyad’ and FB/F stands for ‘feedback/forward’. Again here branching is dominant, but with a stronger preference for merging and chaining. The order of the patterns (i.e. {Branch, Merge, Chain}) does not change, but in the author network branching is not as dominant as in the content network and in fact merging and chaining are not only frequent, but also significantly so. This indicates that authors exhibit a strong tendency to remix from more than one author, as well as to form chains, not so much from remixing the same piece over and over again (see earlier discussion on the low frequency of long chains in the content network), but more often from the remixing of different pieces by the same authors. We thus observe a strong crosspollination across authors through the practice of creative reuse, as well as an indirect cascading of cultural influences from one author to the next.

In contrast to the content network, it is also possible for reciprocal (bidirectional) links to be formed in the author network. A mutual dyad (MD) is formed when two authors reuse each other’s works. We see in Table 6 that 5 significant reuse patterns with reciprocal ties occur with frequencies higher than those of similar random networks. Their absolute frequencies are low, so even in the author network reciprocity is not so common, most authors tend to either focus on remixing the content of others (remixers), or on seeding the community with new content (seeders). However, almost 10% of the patterns in the author network contain at least one reciprocal tie (not all shown in Table 6, as some are neither significant nor very frequent).

Table 6: Author network motifs

Motifs	Author Network		No Contests	
	Frequency	Z-score	Frequency	Z-score
Branch ^a	64.20% ^b	4.3	40.27% ^b	4.0
Merge ^a	19.33% ^b	3.7	32.04% ^b	2.9
Chain ^a	9.81% ^b	2.9	17.18% ^b	1.7
Branch-1MD	0.39% ^b	3.8	0.55% ^b	4.8
Merge-1MD	0.11% ^b	7.2	0.22% ^b	5.7
FB/F-2MD	0.08% ^b	6.8	0.11% ^b	5.0
FB/F-1MD	0.07% ^b	2.0	0.12% ^b	0.4
3MD	0.02% ^b	63.5	0.01% ^b	54.8

^a Most frequently occurring patterns

^b Significant motifs ($p < 0.05$) for patterns with frequency $\geq 0.01\%$

Moreover, the pattern with the most significant difference in frequency compared to a random network (z-score of 63.5) is the one with 3 mutual dyads. This perfectly reciprocal pattern is by no means the most common in the network. It accounts for only 0.02% of 3-node patterns found in ccMixer, but the uniquely high z-score signals that authors in the network form such perfectly reciprocal subnetworks much more frequently than one would expect given all other structural characteristics of the network. Such subnetworks and most other patterns exhibiting reciprocal ties are formed in the strongly connected core of the remix network (see Figure 5-D and Table 4). It is thus the authors in the core of the remix network that exhibit an unusually high preference for creating reciprocal ties. This provides additional support for our earlier finding that the core consists of authors who regard themselves as more or less equal peers, as opposed to the authors in the other components of the network, who are mainly seeders and remixers. Reciprocal reuse ties probably also serve to strengthen the social ties between authors in the core.

When we take contests out of the picture, the frequency of branching is again greatly reduced, as expected. However, unlike in the content network, branching is not only reduced, but in fact becomes a minority pattern, with merging and chaining together accounting now for almost 50% of all 3-node patterns. This provides very strong empirical support for our assumption that many of the authors who participate in contests are one-time contributors and otherwise perhaps of little relevance to the rest of the community. The order of the patterns or the significance of the motifs do not change (for the most part) in the network that remains after removing contests and all their derivatives. So none of the patterns or network motifs can be attributed directly to contests, but in the absence of contests, an ad-hoc remixing community appears to engage much more in the synthesis and cascading of musical ideas from one generation to the next.

6. COMMUNITY DYNAMICS

We have collected data on almost all aspects of community activity for a period of 3.5 years. This allows us to study many aspects of community growth and member activity. However, our intent here is not to produce a complete analysis of the dynamics of the community, but to specifically try to address our concerns with respect to the value of contests. Table 7 provides a summary of all contests that have taken place in ccMixer. Each contest has a timeframe defined by a start date and a deadline. The deadline is

sometimes extended. Where records of such extensions have been kept, we have included them in the “Deadline” column.

Table 7: History of ccMixer contests

Contests	Start Date	Deadline	Sources	Entries	Total
Freestyle Mix	15-Dec-04	15-Feb-05	24	77	101
Militia Mix	15-Dec-04	15-Feb-05	4	53	57
Magnatune	01-Jul-05	31-Jul-05	0 ^a	161	161
Copyright Criminals	06-Dec-05	14-Mar-06	1	104	105
Fort Minor	22-Mar-06	6-May-06	2	446	448
Crammed	26-Apr-06	24-May-06	6	96	102
Ghostly	29-Nov-06	27-Dec-06	3	74	77
Vieux	7-Feb-07	7-Mar-07	2	67	69
DJ Vadim	28-Mar-07	19-Apr-07	5	195	200
Curve	6-Jun-07	28-Jun-07	6	51	57
Salman	6-Jul-07	31-Jul-07	4	30	34
Total			57	1354	1411

^a The source material for this contest was pooled from another website

Table 7 also shows the numbers of user-submitted entries per contest, while the numbers of source files handed to community members to remix per contest are reported separately, under “Sources”. We did not make this distinction earlier in this paper, in Table 1, but now it is useful, as we want to see how much attention different contests were able to garner. Also, in Table 1 we did not include the entries of the Magnatune contest (third row of Table 7), because this was a contest where ccMixer members were asked to remix content hosted by the Magnatune music label on a separate website.

We notice that all contests generate a fair amount of remixing activity, although some contests stand out: Fort Minor has been the most successful contest, in terms of remixing activity, with 446 entries, with the DJ Vadim contest being a distant second with 195 entries, followed by the Magnatune contest at 161 entries. Figure 8 shows the history of daily registrations and uploads to the community, from its inception and until January 2008. Cumulative values are overlaid on the secondary axis. We notice several spikes in daily activity, which, when combined with the dates in Table 7, coincide in many occasions with contests. Specifically, spikes tend to appear closer to the deadline of a contest, as more new members sign up and post their remixes for consideration in the contest. So it would appear that contests not only drive some of the remixing activity in the network, but also serve as magnets for the attraction of new members.

We also notice that the largest spike in Figure 8 (Mar-May 2006) coincides with the Fort Minor contest, which was apparently successful not only in generating more remixing activity within the community, but also in attracting many new members. It is even more interesting to note that the growth rate of the community more than doubled at the time when the Fort Minor contest started, in March of 2006, from an average of 6 new registrations per day to 14, and that the higher growth rate was sustained long after the contest had ended, even during periods of no contest activity. Remixing activity also more than doubled, from an average of 4.3 uploads per day, to a new rate of 10.2 uploads per day, which again has remained relatively stable since.

It appears as though the community was able to make a transition to a new phase of sustained higher growth at the same time as the Fort Minor contest. The evidence presented is not sufficient to conclude that there exists a causal relationship between the Fort Minor contest and the long-term increase in the growth rate, but it does motivate a new hypothesis for further research: it may be that contests which attain a certain threshold of publicity can propel a community to a higher plateau of growth, in spite of the fact that they generate a lot of short-lived activity.

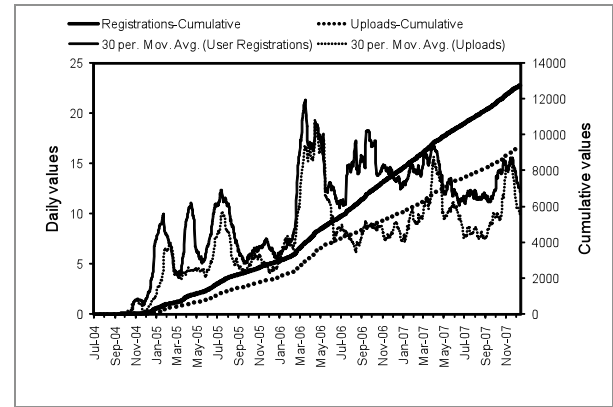


Figure 8: History of registrations and uploads

7. CONCLUSION AND FUTURE WORK

We have examined the network structure of the ccMixer online music community and the effects of contests on the structure and dynamics of remixing activity in the community. For this we introduced the concepts of a reuse network, reuse breadth and reuse depth. We showed that both reuse depth and breadth are limited by cultural constraints that act against the reuse of highly derivative works. Reuse is thus somewhat limited in its ability to continuously generate new works from the same source material. However, we showed that reuse can at least double the output of a community as well as increase the diversity of the output through a process of crosspollination between authors and their works.

The organization of contests helps attract new members to the community and it also leads to short-term spikes in productivity. However, many of the authors introduced to the community through contests do not contribute beyond the duration of the contest. Such authors remain at the periphery of the social structure of the community. On the other hand, the community exhibits a preference for the creation of strong, reciprocal ties between authors who treat each other as peers, thus forming a small core that is the most productive and likely also essential for the sustainability of the community. Contests have virtually no impact on this core. It thus appears that even if community members may not be necessarily averse to the introduction of extrinsic incentives for participation, extrinsic motivation is limited in its ability to grow the core of the community. Perhaps it does serve a useful purpose though, in that it helps grow the audience for all contributing authors, enabling the community to reach a certain critical mass through wider publicity.

One future avenue for this work is to locate our findings from this paper in an analysis of ccMixer as a Community of Practice (CoP). CoPs are groups of people who are brought together by

shared goals and strive to achieve those goals through a process of social learning and membership. ccMixer is quite clearly a CoP in the fact that the members share the common goal of open sharing and reuse of music on the Internet, and they become members of this community through learning and engaging in the practice of remixing. We intend to interview both core and peripheral members of the ccMixer community in order to better understand how they identify with the community and what motivates them to contribute. By doing so, we can more clearly understand who remixes music, how and why it is done.

Another avenue will be to examine the patterns of reuse brought out in this paper in light of the literature on information diffusion in online environments. It will also be interesting to explore differences across communities working with different media types, as we believe that the structure and dynamics of remixing will be different, depending on the media being reused in these communities. Additionally, we also envision comparing and contrasting the creation and reuse of products of cultural or entertainment value to that of products of mainly functional value, such as software. One key difference in objectives would be that for goods of functional value individual components are typically built within larger projects and must conform to a specification such that they may work together, whereas in cultural production there is often no such requirement.

8. ACKNOWLEDGMENTS

We thank Mike Linksvayer and Victor Stone for being instrumental with the data collection. We also wish to thank Ankit Guglani, Clint Gono, Lada Adamic and Giri Kumar Tayi for their valuable inputs. We are also grateful to Lonce Wyse for helping instigate this collaboration.

9. REFERENCES

- [1] Adamic, L. A., Zhang, J., Bakshy, E., and Ackerman, M. S. 2008. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In Proceedings of the 17th international Conference on World Wide Web (Beijing, China, April 21 - 25, 2008). WWW '08. ACM, New York, NY, 665-674.
- [2] Anderson, C. 2006. *The Long Tail: Why the Future of Business is Selling Less of More*. Hyperion, London, UK.
- [3] Bagatelj, V. and Mrvar, A. 1998. PAJEK - Program for Large Network Analysis. *Connections* 21:47-57.
- [4] Bauwens, M. 2005. The Political Economy of Peer Production. Retrieved on Nov 23, 2008 from <http://www.ctheory.net/articles.aspx?id=499>
- [5] Benkler, Y. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, New Haven.
- [6] Borgatti, S.P. 2002. NetDraw: Graph Visualization Software. Analytic Technologies, Harvard.
- [7] Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. and Wiener, J. 2000. Graph Structure In The Web. *Computer Networks* 33(1-6):309-320.
- [8] Cheliotis, G., Chik, W., Guglani, A. & Tayi, G. 2008. Taking Stock of the Creative Commons Experiment. In Proceedings of the 35th Research Conference on Communication, Information and Internet Policy (TPRC) (Arlington VA, Sep 28-30, 2008). Retrieved on Dec 10, 2008 from <http://web.si.umich.edu/tprc/papers/2007/805/CreateCommExp.pdf>
- [9] Clauset, A., Shalizi, C. R., and Newman, M. E. J. 2007. Power-law Distributions in Empirical Data. Retrieved on Dec 10, 2008 from <http://arxiv.org/abs/0706.1062>
- [10] Gosh, R.A. (ed.). 2006. *CODE: Collaborative Ownership and the Digital Economy*. MIT Press, Cambridge.
- [11] Grossman, L. 'Time's Person of the Year: You'. *Time Magazine*, 13 Dec. 2006. Retrieved on Nov 23, 2008 from <http://www.time.com/time/magazine/article/0,9171,1569514,00.html>
- [12] Lessig, L. 2004. *Free Culture: The Nature and Future of Creativity*. Penguin, London, UK.
- [13] Lessig, L. 2008. *Remix: Making Art and Commerce Thrive in the Hybrid Economy*. Penguin, New York.
- [14] Manovich, L. 2007. Cultural Analytics: Analysis and Visualization of Large Cultural Datasets. Retrieved on Nov 23, 2008 from http://www.manovich.net/cultural_analytics.pdf
- [15] Manovich, L. 2005. Remixing and Remixability. Retrieved on Jan 10, 2008 from http://www.manovich.net/DOCS/Remix_modular.doc
- [16] Miller, P.D. 2004. *Rhythm Science*. MIT Press, Cambridge.
- [17] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. 2002. Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594):824-827.
- [18] Mitzenmacher, M. 2003. A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics* 1(2): 226-251.
- [19] Newman, M., Barabasi, A.-L., and Watts, D. (eds.). 2006. *The Structure and Dynamics of Networks*. Princeton Studies in Complexity, Princeton University Press, Princeton.
- [20] Vickery, G. and Wunsch-Vincent, S. 2007. *Participative Web and User-Created Content: Web 2.0 Wikis and Social Networking*. OECD e-Publication.
- [21] Von Hippel, E. 2006. *Democratizing Innovation*. MIT Press, Cambridge.
- [22] Wasserman, S. and Faust, K. 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.
- [23] Wernicke, S. and Rasche, F. 2006. FANMOD: A Tool for Fast Network Motif Detection. *Bioinformatics* 22(9):1152-1153.
- [24] Zhang, J., Ackerman, M. S., and Adamic, L. 2007. Expertise networks in online communities: structure and algorithms. In Proceedings of the 16th international Conference on World Wide Web (Banff, Alberta, Canada, May 08 - 12, 2007). WWW '07. ACM, New York, NY, 221-230.