Collective dynamics of social annotation

Ciro Cattuto[a], Alain Barrat[a,b], Andrea Baldassarri[c], Gregory Schehr[d], and Vittorio Loreto[a,c,1]

[a]Complex Networks Lagrange Laboratory, Institute for Scientific Interchange, 10133 Turin, Italy; [b]Centre de Physique Théorique, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 6207, Campus de Luminy, Case 907, 13288 Marseille Cedex 9, France; [c]Dipartimento di Fisica, ''Sapienza'' Università di Roma, Piazzale Aldo Moro 5, 00185 Rome, Italy; and [d]Laboratoire de Physique Théorique Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8627, Université Paris-Sud, 91405 Orsay Cedex, France

Edited by H. Eugene Stanley, Boston University, Boston, MA, and approved April 17, 2009 (received for review February 2, 2009)

The enormous increase of popularity and use of the worldwide web has led in the recent years to important changes in the ways people communicate. An interesting example of this fact is provided by the now very popular social annotation systems, through which users annotate resources (such as web pages or digital photographs) with keywords known as ''tags.'' Understanding the rich emergent structures resulting from the uncoordinated actions of users calls for an interdisciplinary effort. In particular concepts borrowed from statistical physics, such as random walks (RWs), and complex networks theory, can effectively contribute to the mathematical modeling of social annotation systems. Here, we show that the process of social annotation can be seen as a collective but uncoordinated exploration of an underlying semantic space, pictured as a graph, through a series of RWs. This modeling framework reproduces several aspects, thus far unexplained, of social annotation, among which are the peculiar growth of the size of the vocabulary used by the community and its complex network structure that represents an externalization of semantic structures grounded in cognition and that are typically hard to access.

networks theory | statistical physics | social web | emergent semantics | web-based systems

The rise of Web 2.0 has dramatically changed the way in which information is stored and accessed and the relationship between information and online users. This is prompting the need for a research agenda about ''web science,'' as put forward in ref. 1. A central role is played by user-driven information networks, i.e., networks of online resources built in a bottom-up fashion by web users. These networks entangle cognitive, behavioral and social aspects of human agents with the structure of the underlying technological system, effectively creating technosocial systems that display rich emergent features and emergent semantics (2, 3). Understanding their structure and evolution brings forth new challenges.

Many popular web applications are now exploiting user-driven information networks built by means of social annotations (4, 5). Social annotations are freely established associations between web resources and metadata (keywords, categories, ratings) performed by a community of web users with little or no central coordination. A mechanism of this kind that has swiftly become well established is that of collaborative tagging (see www.adammathes.com/academic/computer-mediated-communication/folksonomies.html) (6), whereby web users associate freeform keywords—called ''tags''—with online content such as web pages, digital photographs, bibliographic references, and other media. The product of the users' tagging activity is an open-ended information network—commonly referred to as ''folksonomy''—which can be used for navigation and recommendation of content and has been the object of many recent investigations across different disciplines (7, 8). Here, we show how simple concepts borrowed from statistical physics and the study of complex networks can provide a modeling framework for the dynamics of collaborative tagging and the structure of the ensuing folksonomy.

Two main aspects of the social annotation process, so far unexplained, deserve special attention. One striking feature is the so-called Heaps' law (9) (also known as Herdan's law in linguistics), originally studied in information retrieval for its relevance for indexing schemes (10). Heaps' law is an empirical law that describes the growth in a text of the number of distinct words as a function of the number of total words scanned. It describes, thus, the rate of innovation in a stream of words, where innovation means the adoption for the first time in the text of a given word. This law, also experimentally observed in streams of tags, consists of a power law with a sublinear behavior (8, 11). In this case, the rate of innovation is the rate of introduction of new tags, and a sublinear behavior corresponds to a rate of adoption of new words or tags decreasing with the total number of words (or tags) scanned. Most existing studies about Heaps' law, either in information retrieval or in linguistics, explained it as a consequence of the so-called Zipf's law (12) [see ref. 10 and supporting information (SI)]. It would instead be highly desirable to have an explanation for it relying only on very basic assumptions on the mechanisms behind social annotation.

Another important way to analyze user-driven information networks is given by the framework of complex networks (13–15). These structures are, indeed, user-driven information networks (16), i.e., networks linking (for instance) online resources, tags, and users, built in a bottom-up fashion through the uncoordinated activity of thousands to millions of web users. We shall focus in particular on the particular structure of the so-called cooccurrence network. The cooccurrence network is a weighted network where nodes are tags, and 2 tags are linked if they were used together by at least 1 user, the weight being larger when this simultaneous use is shared by many users. Correlations between tag occurrences are (at least partially) an externalization of the relations between the corresponding meanings (17, 18) and have been used to infer formal representations of knowledge from social annotations (19). Notice that cooccurrence of 2 tags is not a priori equivalent to a semantic link between the meanings/concepts associated with those tags and that understanding what cooccurrence precisely means, in terms of semantic relations of the cooccurring tags, is an open question that is investigated in more applied contexts (20, 21).

On these aspects of social annotation systems, a certain number of stylized facts about, e.g., tag frequencies (6, 8) or the growth of the tag vocabulary (11), have been reported, but no modeling framework exists that can naturally account for them while reproducing the cooccurrence network structure. Here, we ask whether the structure of the cooccurrence network can be explained in terms of a generative model and how the structure of the experimentally observed cooccurrence network is related

Author contributions: C.C., A. Barrat, A. Baldassarri, G.S., and V.L. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: vittorio.loreto@roma1.infn.it.

This article contains supporting information online at www.pnas.org/cgi/content/full/0901136106/DCSupplemental.

to the underlying hypotheses of the modeling scheme. We show in particular that the idea of social exploration of a semantic space has more than a metaphorical value and actually allows us to reproduce simultaneously a set of independent correlations and fine observables of tag cooccurrence networks as well as robust stylized facts of collaborative tagging systems.

User-Driven Information Networks. We investigate user-driven information networks using data from 2 social bookmarking systems: del.icio.us† and BibSonomy.‡ Del.icio.us is a very popular system for bookmarking web pages and pioneered the mechanisms of collaborative tagging. It hosts a large body of social annotations that have been used for several scientific investigations. BibSonomy is a smaller system for bookmarking bibliographic references and web pages (22). Both del.icio.us and BibSonomy are broad folksonomies (see www.personalinfocloud.com/2005/02/), in which users provide metadata about preexisting resources and multiple annotations are possible for the same resource, making the ensuing tagging patterns truly "social" and allowing their statistical characterization. A more detailed description of the datasets is given in the SI.

A single user annotation, also known as a post, is a triple of the form (u, r, T), where u is a user identification, r is the unique identification of a resource (a URL pointing to a web page, for the systems under study), and $T = \{t_1, t_2, \dots\}$ is a set of tags represented as text strings. We define the tag cooccurrence network based on post cooccurrence. That is, given a set of posts, we create an undirected and weighted network where nodes are tags and 2 tags, t_1 and t_2 are connected by an edge if and only if there exists 1 post in which they were used in conjunction. The weight $w_{t_1t_2}$ of an edge between tags t_1 and t_2 can be naturally defined as the number of distinct posts where t_1 and t_2 cooccur. This construction reflects the existence of semantic correlations between tags and translates the fact that these correlations are stronger between tags cooccurring more frequently. We emphasize once again that the cooccurrence network is an externalization of hidden semantic links and therefore distinct from underlying semantic lexicons or networks.

The study of the global properties of the tagging system, and in particular of the global cooccurrence network, is of interest but mixes potentially many different phenomena. We therefore consider a narrower semantic context, defined as the set of posts containing 1 given tag. We define the vocabulary associated with a given tag t^* as the set of all tags occurring in a post together with t^*, and the time is counted as the number of posts in which t^* has appeared. The size of the vocabulary follows a sublinear power-law growth (Fig. 1), similar to the Heaps' law (9) observed for the vocabulary associated with a given resource and for the global vocabulary (11). Fig. 1 also displays the main properties of the cooccurrence network, as measured by the quantities customarily used to characterize statistically complex networks and to validate models (14, 15). These quantities can be separated in 2 groups. On the one hand, they include the distributions of single node or single link quantities, whose investigations allow one to distinguish between homogeneous and heterogeneous systems. Fig. 1 shows that the cooccurrence networks display broad distributions of node degrees k_t (number of neighbors of node t), node strengths s_t (sum of the weights of the links connected to t, $s_t = \Sigma_{t'} w_{tt'}$), and link weights. The average strength $s(k)$ of vertices with degree k, $s(k) = 1/N_k \Sigma_{t/k_t=k} s_t$, where N_k is the number of nodes of degree k, also shows that correlations between topological information and weights are present. On the other hand, these distributions by themselves are not sufficient to fully characterize a network, and higher-order

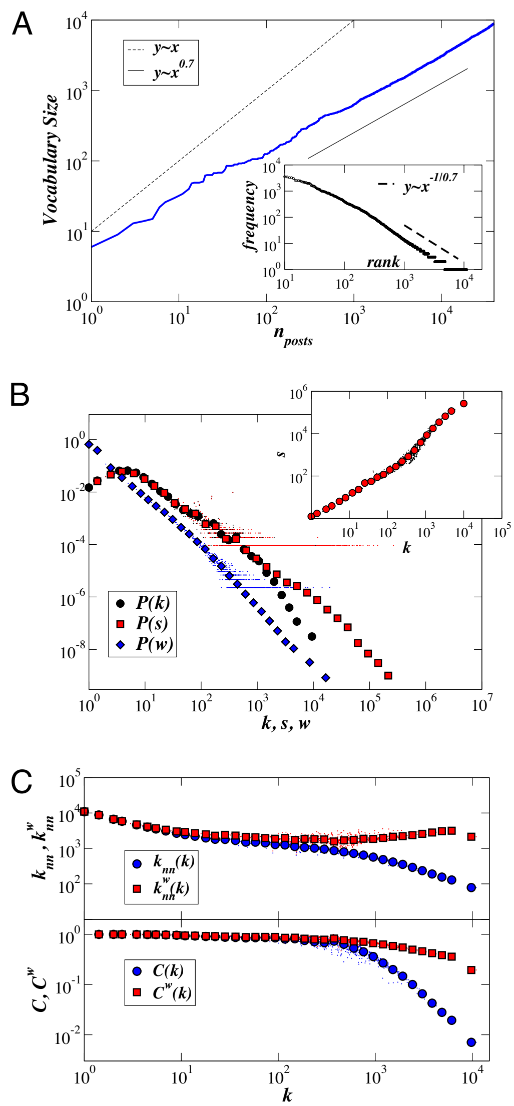

Fig. 1. Data corresponding to the posts containing the tag *Folksonomy* in del.icio.us. (*A*) Heaps' law: growth of the vocabulary size associated with the tag t^* = *Folksonomy*, measured as the number of distinct tags cooccurring with t^*, as a function of the number n_{posts} of posts containing t^*. The dotted line corresponds to a linear growth law, whereas the continuous line is a power-law growth with exponent 0.7. (*Inset*) Frequency-rank plot for the tags. The dashed line corresponds to a power law $-1.42 \simeq -1./0.7$. (*B* and *C*) Main properties of the cooccurrence network of the tags cooccurring with the tag *Folksonomy* in del.icio.us, built as described in *User-Driven Information Networks*. (*B*) Broad distributions of degrees k, strengths s and weights w are observed. The *Inset* shows the average strength of nodes of degree k, with a superlinear growth at large k. (*C*) Weighted (k^w_{nn}) and unweighted (k_{nn}) average degree of nearest neighbors (*Upper*) and weighted (C^w) and unweighted (C) (*Lower*) average clustering coefficients of nodes of degree k. k_{nn} displays a disassortative trend, and a strong clustering is observed. At small k, the weights are close to 1 ($s(k) \approx k$, see *B Inset*), and $k^w_{nn} \approx k_{nn}$, $C^w \approx C$. At large k instead, $k^w_{nn} > k_{nn}$ and $C^w > C$, showing that large weights are preferentially connecting nodes with large degree: Large-degree nodes are joined by links of large weight, i.e., they cooccur frequently. In *B* and *C* both raw and logarithmically binned data are shown.

correlations have to be investigated. In particular, the average nearest-neighbor degree of a vertex t, $k_{nn,t} = 1/k_t \Sigma_{t' \in \mathcal{V}(t)} k_{t'}$, where $\mathcal{V}(t)$ is the set of t's neighbors, gives information on correlations between the degrees of neighboring nodes. Moreover, the clustering coefficient $c_t = e_t/(k_t(k_t - 1)/2)$ of a node t measures local cohesiveness through the ratio between the number e_t of links

†http://del.icio.us

‡http://www.bibsonomy.org

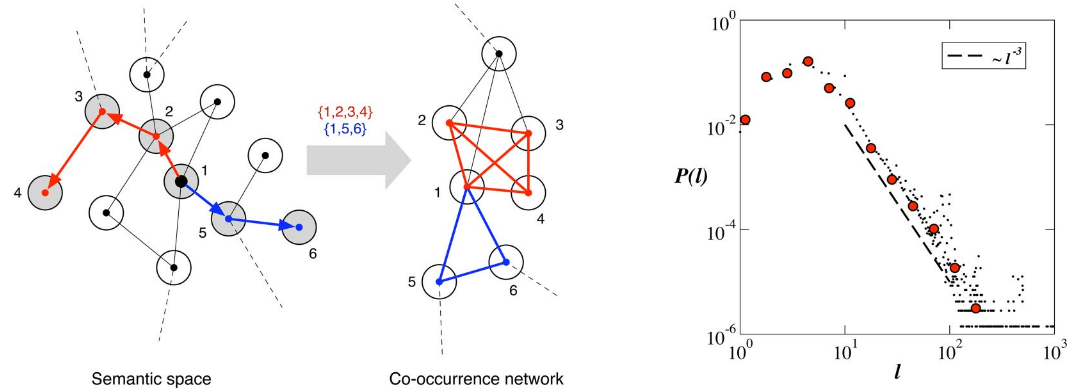

Fig. 2. (*Left*) Illustration of the proposed mechanism of social annotation. The semantic space is pictured as a network in which nodes represent tags, and a link corresponds to the possibility of a semantic association between tags. A post is then represented as a RW on the network. Successive RWs starting from the same node allow the exploration of the network associated with a tag (here pictured as node 1). The artificial cooccurrence network is built by creating a clique between all nodes visited by a RW. (*Right*) Empirical distribution of posts' lengths $P(l)$. A power-law decay $\approx l^{-3}$ (dashed line) is observed.

between the k_t neighbors of t and the maximum number of such links (23). The functions $k_{nn}(k) = 1/N_k \Sigma_{t/k_t=k} k_{nn,t}$ and $C(k) = 1/N_k \Sigma_{t/k_t=k} c_t$ are convenient summaries of these quantities, that can also be generalized to include weights [see SI for the definitions of $k^w_{nn}(k)$ and $C^w(k)$]. Fig. 1 shows that broad distributions and nontrivial correlations are observed. All of the measured features are robust across tags within 1 tagging system and across the tagging systems we investigated (see SI).

Modeling Social Annotation. The observed features are emergent characteristics of the uncoordinated action of a user community, which call for a rationalization and for a modeling framework. We now present a simple mechanism able to reproduce the complex evolution and structure of the empirical data.

The fundamental idea underlying our approach, illustrated in Fig. 2, is that a post corresponds to a random walk (RW) of the user in a "semantic space" modeled as a graph. Starting from a given tag, the user adds other tags, going from 1 tag to another by semantic association. It is then natural to picture the semantic space as network-like, with nodes representing tags, and links representing the possibility of a semantic link (24). A precise and complete description of such a semantic network being out of reach, we make very general hypothesis about its structure and we have checked the robustness of our results with respect to different plausible choices of the graph structure (24). Nevertheless, as we shall see later on, our results help fixing some constraints on the structural properties of such a semantic space: it should have a finite average degree together with a small graph diameter, which ensures that RWs starting from a fixed node and of limited length can potentially reach all nodes of the graph. In this framework, the vocabulary cooccurring with a tag is associated with the ensemble of nodes reached by successive RWs starting from a given node, and its size with the number of distinct visited nodes, N_{distinct}, which grows as a function of the number n_{RW} of performed RWs. Empirical evidence on the distribution of post lengths (Fig. 2) suggests that one consider RWs of random lengths, distributed according to a broad law (see SI for the case of walks of fixed length). Analytical and numerical investigations show that sublinear power law-like growths of N_{distinct} are then generically observed, mimicking the Heaps' law observed in tagging systems (Fig. 3 and SI).

Synthetic Cooccurrence Networks. Vocabulary growth is only one aspect of the dynamics of tagging systems. Networks of cooccurrence carry much more information and exhibit very specific features (Fig. 1). Our approach allows one to construct synthetic cooccurrence networks: We associate to each RW a clique formed by the nodes visited (see Fig. 2) and consider the union of the n_{RW} such cliques. Moreover, each link i–j built in this way receives a weight equal to the number of times nodes i and j appear together in a RW. This construction mimics precisely the construction of the empirical cooccurrence network and reflects the idea that tags that are far apart in the underlying semantic network are visited together less often than tags that are semantically closer. Figs. 3 and 4 show how the synthetic networks reproduce all statistical characteristics of the empirical data (Fig. 1), both topological and weighted, including highly nontrivial correlations between topology and weights. Fig. 4 in particular explores how the weight w_{ij} of a link is correlated with its extremities' degrees k_i and k_j. The peculiar shape of the curve can be understood within our framework. First, the broad distribution in l is responsible for the plateau ≈ 1 at small values of $k_i k_j$, because it corresponds to long RWs that occur rarely and visit nodes that will be typically reached a very small number of times (hence small weights). Moreover, $w_{ij} \approx (k_i k_j)^a$ at large weights. Denoting by f_i the number of times node i is visited, $w_{ij} \approx f_i f_j$ in a mean-field approximation that neglects correlations. On the other hand, k_i is by definition the number of distinct nodes visited together with node i. Restricting the RWs to the only processes that visit i, it is reasonable to assume that such sampling preserves Heaps' law, so that $k_i \propto f_i^{\alpha}$, where α is the growth exponent for the global process. This leads to $w_{ij} \approx (k_i k_j)^a$ with $a = 1/\alpha$. Because $\alpha \simeq 0.7$–0.8, we obtain a close to 1.3–1.5, consistently with the numerics.

Strikingly, the synthetic cooccurrence networks reproduce also other, more subtle observables, such as the distribution of cosine similarities between nodes. In a weighted network, the similarity of 2 nodes i_1 and i_2 can be defined as

$$\mathrm{sim}(i_1, i_2) \equiv \frac{\sum_j w_{i_1 j} w_{i_2 j}}{\sqrt{\sum_\ell w^2_{i_1 \ell} \sum_\ell w^2_{i_2 \ell}}}, \qquad [1]$$

which is the scalar product of the vectors of normalized weights of nodes i_1 and i_2. This quantity, which measures the similarities between neighborhoods of nodes, contains semantic information that can be used to detect synonymy relations between tags or to uncover "concepts" from social annotations (20). Fig. 5 shows the histograms of pairwise similarities between nodes in real and synthetic cooccurrence networks. The distributions are very similar, with a skewed behavior and a peak for low values of the similarities. In the SI, we report the similarity distributions for other tags and provide a more detailed discussion on their properties.

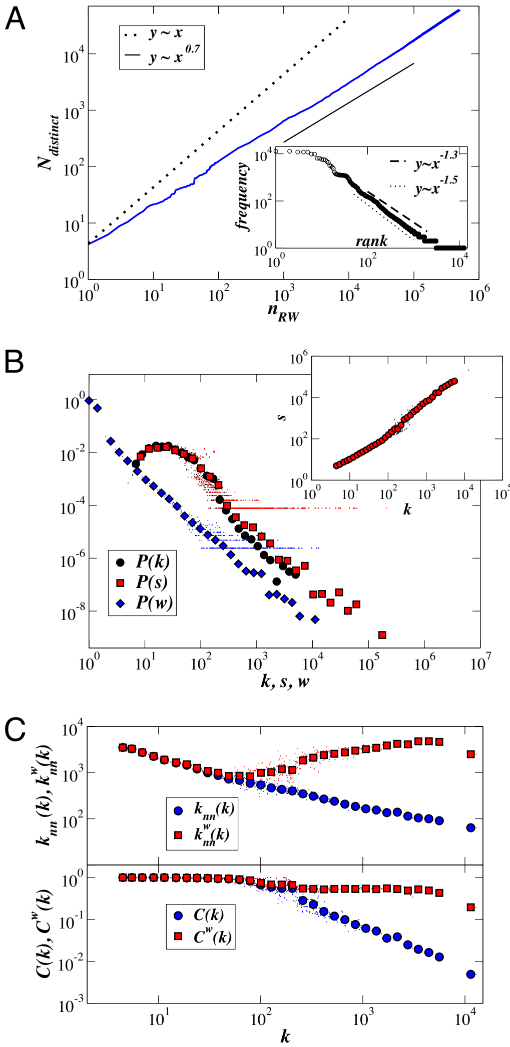

Fig. 3. Synthetic data produced through the proposed mechanism. (*A*) Growth of the number of distinct visited sites as a function of the number of RWs performed on a Watts–Strogatz network (see SI) of size 5×10^4 nodes and average degree 8, rewiring probability $P = 0.1$. Each RW has a random length l taken from a distribution $P(l) \approx l^{-3}$. The dotted line corresponds to a linear growth law, whereas the continuous line is a power-law growth with exponent 0.7. (*Inset*) Frequency-rank plot. The dashed and dotted line have slope -1.3 and -1.5, respectively. (*B* and *C*) Properties of the synthetic cooccurrence network obtained for $n_{\mathrm{RW}} = 5 \times 10^4$, to be compared with the empirical data of Fig. 1.

Whereas the data shown in Figs. 3 and 4 correspond to a particular example of underlying network (a Watts–Strogatz network, see ref. 23 and SI), taken as a sketch for the semantic space, we investigate in the SI the dependence of the synthetic network properties on the structure of the semantic space and on the other parameters, such as n_{RW} or the distribution of the RW lengths. Interestingly, we find an overall extremely robust behavior for the diverse synthetic networks, showing that the proposed mechanism reproduces the empirical data without any need for strong hypotheses on the semantic space structure. The only general constraints implied by the mechanism proposed here are the existence of an underlying semantic graph with a small diameter and a finite average degree (RWs on a fully connected graph would not work, for instance) and a broad distribution of post lengths. This lack of strong constraints on the precise structure of the underlying semantic network is actually

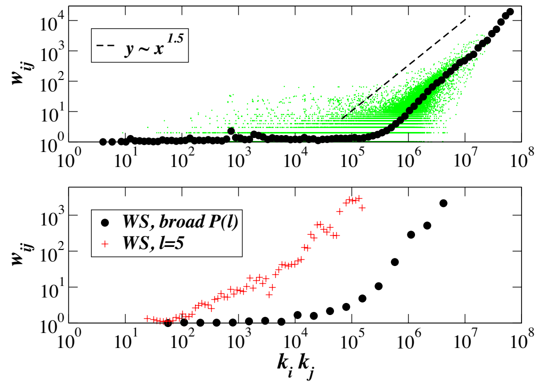

Fig. 4. Correlations between the weights of the links in the cooccurrence networks and the degrees of the links' endpoints, as measured by plotting the weight w_{ij} of a link i–j versus the product of the degrees $k_i k_j$. (*Upper*) Cooccurrence network of the tag Folksonomy of del.icio.us; each green dot corresponds to a link; the black circles represent the average over all links i–j with given product $k_i k_j$. (*Lower*) Synthetic cooccurrence networks obtained from $n_{\mathrm{RW}} = 5 \times 10^4$ RWs performed on a Watts–Strogatz network of 10^5 nodes. The black circles correspond to RWs of random lengths distributed according to $P(l) \approx l^{-3}$ and the red crosses to fixed-length RWs ($l = 5$).

a remarkable feature of the proposed mechanism. The details of the underlying network will unavoidably depend on the context, namely on the specific choice of the central tag t^*, and the robustness of the generative model matches the robustness of the features observed in cooccurrence networks from real systems. Of course, given an empiric cooccurrence network, a careful simultaneous fitting procedure of the various observables would be needed to choose the most general class of semantic network structures that generate that specific network by means of the mechanism introduced here. This delicate issue goes beyond the goal of this article and raises the open question of the definition of the minimal set of statistical observables needed to specify a graph (25).

Conclusions

Investigating the interplay of human and technological factors in user-driven systems is crucial to understand the evolution and the potential impact these systems will have on our societies. Here, we have shown that sophisticated features of the information networks stemming from social annotations can be

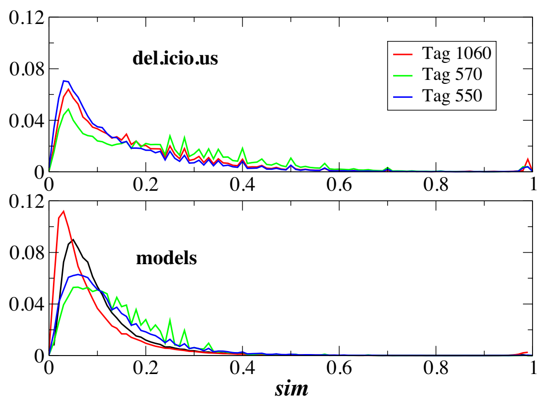

Fig. 5. Distribution of tag similarities for real (*Upper*) and synthetic (*Lower*) cooccurrence networks. For del.icio.us, the tag number represents its popularity rank in the database. For the synthetic cooccurrence networks, the different curves correspond to different underlying networks on which the RWs are performed (see SI).

captured by regarding the process of social annotation as a collective exploration of a semantic space, modeled as a graph, by means of a series of RWs. The proposed generative mechanism naturally yields an explanation for the Heaps' law observed for the growth of tag vocabularies. The properties of the cooccurrence networks generated by this mechanism are robust with respect to the details of the underlying graph, provided it has a small diameter and a small average degree. This mirrors the robustness of the stylized facts observed in the experimental data, across different systems.

Networks of resources, users, and metadata such as tags have become a central collective artifact of the information society. These networks expose aspects of semantics and of human dynamics, and are situated at the core of innovative applications. Because of their novelty, research about their structure and evolution has been mostly confined to applicative contexts. The results presented here are a definite step toward a fundamental understanding of user-driven information networks that can prompt interesting developments, because they involve the application of recently developed tools from complex networks theory to this new domain. An open problem, for instance, is the generalization of our modeling approach to the case of the full hypergraph of social annotations, of which the cooccurrence network is a projection. Moreover, user-driven information networks lend themselves to the investigation of the interplay between social behavior and semantics, with theoretical and applicative outcomes such as node ranking (i.e., for search and recommendation), detection of nonsocial behavior (such as spam), and the development of algorithms to learn semantic relations from a large-scale dataset of social annotations.

ACKNOWLEDGMENTS. We thank A. Capocci, H. Hilhorst, and V. D. P. Servedio for many interesting discussions and suggestions. This research has been partly supported by the TAGora project funded by the Future and Emerging Technologies program of the European Commission under Contract IST-34721. V.L. is part of the research network A Topological Approach To Cultural Dynamics, supported by the Sixth Framework Programme of the European Union (project no. 043415).

1. Berners-Lee T, Hall W, Hendler J, Shadbolt N, Weitzner DJ (2006) Creating a Science of the Web. *Science* 313:769–771.
2. Staab S, Santini S, Nack F, Steels L, Maedche A (2002) Emergent semantics. *Intel Syst IEEE* [see also *IEEE Expert*] 17:78–86.
3. Mika P (2007) Ontologies are us: A unified model of social networks and semantics. *Web Semant* 5:5–15.
4. Wu X, Zhang L, Yu Y (2006) Exploring social annotations for the semantic web. (Assoc for Comput Machinery, New York), pp 417–426.
5. Hammond T, Hannay T, Lund B, Scott J (2005) Social bookmarking tools (I): A general review. *D-Lib Mag*, doi:10.1045/april2005-hammond.
6. Golder S, Huberman BA (2006) The structure of collaborative tagging systems. *J Info Sci* 32:198–208.
7. Marlow C, Naaman M, Boyd D, Davis M (2006) *HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, To Read* (Assoc Comput Machinery, New York), pp 31–40.
8. Cattuto C, Loreto V, Pietronero L (2007) Semiotic dynamics and collaborative tagging. *Proc Natl Acad Sci USA* 104:1461–1464.
9. Heaps HS (1978) *Information Retrieval: Computational and Theoretical Aspects.* (Academic, Orlando, FL).
10. Baeza-Yates RA, Navarro G (2000) Block addressing indices for approximate text retrieval. *J Am Soc Info Sci* 51:69–82.
11. Cattuto C, Baldassarri A, Servedio VDP, Loreto V (2007) Vocabulary growth in collaborative tagging systems. http://arxiv.org/abs/0704.3316.
12. Zipf GK (1949) *Human Behavior and the Principle of Least Effort* (Addison–Wesley, Reading, MA).
13. Dorogovtsev S, Mendes J (2003) *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford Univ Press, Oxford).
14. Pastor-Satorras R, Vespignani A (2004) *Evolution and Structure of the Internet: A Statistical Physics Approach.* (Cambridge Univ Press, Cambridge, UK).
15. Barrat A, Barthélemy M, Vespignani A (2008) *Dynamical Processes on Complex Networks.* (Cambridge Univ Press, Cambridge, UK).
16. Cattuto C, et al. (2007) Network properties of folksonomies. *AI Commun J* (Special Issue on Network Analysis in Natural Sciences and Engineering) 20:245–262.
17. Sowa JF (1984) *Conceptual Structures: Information Processing in Mind and Machine.* (Addison–Wesley Longman, Boston).
18. Sole RV, Corominas B, Valverde S, Steels L (2005) Language networks: Their structure, function and evolution. Working Paper 05-12-042 (Santa Fe Institute, Santa Fe, NM).
19. Heymann P, Garcia-Molina H (2006) *Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems*, Tech Rep 2006-10.
20. Cattuto C, Benz D, Hotho A, Stumme G (2008) Semantic grounding of tag relatedness in social bookmarking systems. *Lecture Notes Comput Sci* 5318:615–631.
21. Cattuto C, Benz D, Hotho A, Stumme G (2008) Semantic analysis of tag similarity measures in collaborative tagging systems in *Proceedings of the 3rd Workshop on Ontology Learning and Population (OLP3)* (Assoc Comput Linguistics, Stroudsburg, PA).
22. Hotho A, Jäschke R, Schmitz C, Stumme G (2006) *BibSonomy: A Social Bookmark and Publication Sharing System*, eds de Moor A, Polovina S, and Delugach H (Aalborg Univ Press, Aalborg, Denmark).
23. Watts DJ, Strogatz SH (1998) Collective dynamics of "small-world" networks. *Nature* 393:440–442.
24. Steyvers M, Tenenbaum JB (2005) The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognit Sci* 29:41–78.
25. Mahadevan P, Krioukov D, Fall K, Vahdat A (2006) Systematic topology analysis and generation using degree correlations in *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (Assoc Comput Machinery, New York), pp 135–146.