
Socio-technical Computation

Markus Luczak-Roesch

University of Southampton
Web and Internet Science
Southampton, SO17 1BJ, UK
mlr1m12@soton.ac.uk

Ramine Tinati

University of Southampton
Web and Internet Science
Southampton, SO17 1BJ, UK
r.tinati@soton.ac.uk

Kieron O'Hara

University of Southampton
Web and Internet Science
Southampton, SO17 1BJ, UK
kmo@ecs.soton.ac.uk

Nigel Shadbolt

University of Southampton
Web and Internet Science
Southampton, SO17 1BJ, UK
nrs@ecs.soton.ac.uk

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

CSCW'15 Companion, March 14–18, 2015, Vancouver, BC, Canada.
ACM 978-1-4503-2946-0/15/03.
<http://dx.doi.org/10.1145/2685553.2698991>

Abstract

Motivated by the significant amount of successful collaborative problem solving activity on the Web, we ask: Can the accumulated information propagation behavior on the Web be conceived as a giant machine, and reasoned about accordingly? In this paper we elaborate a thesis about the computational capability embodied in information sharing activities that happen on the Web, which we term socio-technical computation, reflecting not only explicitly conditional activities but also the organic potential residing in information on the Web.

Author Keywords

computational theory; information cascades.

ACM Classification Keywords

F.1.1. **[Models of Computation]**: Bounded-action devices; H.1.2. **[User/Machine Systems]**: Human factors.

Introduction

We observe many successful collaborative problem solving activities on the World Wide Web today, including the collection of donations for charities, the coordination of mass movements and demonstrations, collaborative content authoring, and even individual requests for help and advice. Such emergent

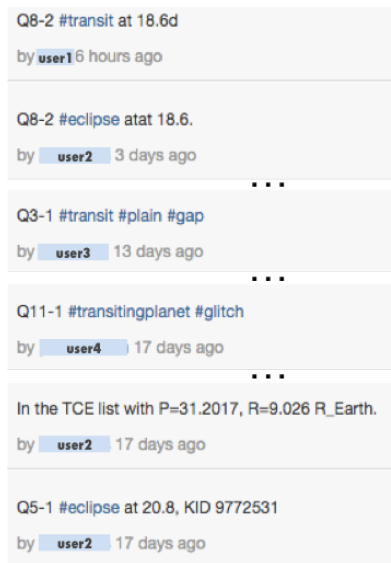


Figure 1. Participants employ self-constructed content patterns for testing hypothesis about objects of interest in the Planet Hunters project on the Zooniverse platform. Hashtags as well as identifiers from remote systems (e.g. the KID refers to an object in a public NASA database) let relationships between posts emerge. The system does not feature any explicit social network but the independent information sharing activities contribute to collective problem solving even spanning across the system boundaries of Planet Hunters.

phenomena are generally characterized by a socio-technical interplay that determines how applications and services on the Web are orchestrated for a particular purpose and how information diffuses.

In many cases explicit social networks built around Web-based systems condition this socio-technical interplay. These can be exploited to determine with increasing accuracy whether a piece of information was published in direct response to another one, which then allows inferences about roles of actors in information diffusion processes for example. However, there are also cases of uncertainty about potential relationships between information, especially when online communities do not feature an explicit social network or system borders are crossed (see Figure 1 for an example from the Planet Hunters¹ project forums on the Zooniverse citizen science platform²). Relationships might be missed out - or appear to be serendipitous - because the trigger event for their creation is not represented within the feature space under investigation (e.g. two people inventing the same tag in independent systems on the Web at almost the same time). This type of example suggests that there exists purposeful collaborative work on the Web that is not necessarily conditioned by binary social links between contributors and therefore does not necessarily leave explicit traces between information that contributes to the higher order goal of the work.

At this point we ask if it is possible to derive a formal model that represents the computational power of human participants publishing information

¹ <http://planethunters.org>

² <http://zooniverse.org>

independently but with a common purpose on the Web. Or in other words: **Does the accumulated information propagation behavior on the Web form a giant machine?**

In this paper we elaborate a thesis about the computational capability embodied in information sharing activities that happen on the Web. We argue that in Web-based systems a specific form of emergent *information cascade* can be seen as abstractly representing computation performed by human users orchestrating the technical capabilities of machines dynamically. Those cascades can be modeled independently of any underlying system-specific or social network features. This work is ultimately targeted at developing a generic model to capture and analyze this organic computational capability of the Web, which we term *socio-technical computation*. It will allow us to equip the design and development of successful problem solving activities on the Web with theoretical underpinnings such as complexity analysis, verification and validation. Additionally, this model will enable the indexing and retrieval of the procedural knowledge on the Web complementing other well-known search functions.

From socially-determined towards transcendental information cascades

It is possible to follow the flow of information through a network, quantifying phenomena such as influence, by observing explicit references to other resources' URIs (e.g. to link to a remote hypertext document or embed a remote image) [1,2,3], and patterns within the content (e.g. a quotation, a meme or a sample in an image) [4]. The viral spread of such patterns can be conceptualized as information flow, information

diffusion, information propagation or *information cascades* and is typically modeled in the following way: one or more undirected sub-networks represent structures of explicit relationships between entities along which information can diffuse (e.g. blog sites interlinked by blogroll features or users forming a following or friendship graph); an actual diffusion process is represented as a time-stamped directed overlay network; each edge in the overlay network is directed from the “infector” node to the “infectee” node as well as labelled with the time when the diffusion was evidenced on the side of the “infectee” and the identifier of the diffusing information; evidence for an infection is inferred based on features of the sub-network. Cascades have a single initiator but they can collide and merge when identifiers from different cascades are used in one node [5,6,7].

We explore the possibility of abstracting the social context away from the technological substrate to understand the Web’s intrinsic information cascades, considering not only local understanding of its use but also an abstract global view. This lets us propose a new model that we call transcendental information cascades. Informed by Kleinberg’s work on burst structures in streams [8] it regards time as the only ascertainable condition for relationships between any two resources and deliberately incorporates serendipity.

Formally the diffusion process is still a directed network as shown by the example in Figure 2. However, we do not presume any sub-network to exist but only a set of Web resources (e.g. individual blog posts, microposts, forum entries, or Web pages) instead. Nodes in the network are those resources from this set that contain one or multiple cascade identifiers. An edge exists

between any two nodes that match the following rule: the two nodes share a unique subset of all identifiers they contain; this subset is not part of the interactions featured by any node that was temporally created between the two.

This cascade model yields different outputs depending on the data to hand determined by the extent of the Web crawl, and the matching algorithm, which determines which cascade identifiers will be spotted (e.g. reuse of hashtags, URIs, quotes, images, or maybe exploiting wider semantics or sentiment). An information cascade in the sense we describe here flows through the Web, channeling and preserving information across time. It therefore has storage and transfer capacity, and as a result is an important aid particularly for distributed communities with few communally-created information storage facilities capable of allowing access to information in a timely manner at the point at which it is needed. Some, but not all, input signals (nodes that use certain identifying pattern for the first time) become output signals (nodes that have no more outgoing edges), so a body of information can evolve over time. Information loss may correspond to information ceasing to be current, or alternatively a cascade might branch to create divergent cascades whose combined capacity may make up for apparent local losses.

Cascade motifs as an indicator of state?

Inspired by this model we define socio-technical computation as the computational capability embodied in cascades of information sharing activities on the Web that are not necessarily conditioned by system-specific or social network features but only time and inherent properties of pairs of resources. The key thesis behind

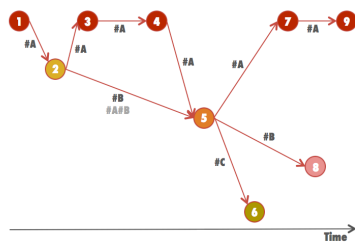


Figure 2. Example cascade using hashtags as identifiers. The visualization represents the sequence “#A” - “#A#B” - “#A” - “#A” - “#A#B#C” - “#C” - “#A” - “#B#D” - “#A”.

this is: Every chronological sequence of explicitly or implicitly related Web resources can be represented by appropriate automata. In other words, we assume that it is possible to model information sharing activity on the Web as cascades that represent the data flow of a computation that can be simulated by state machines.

Under the assumption that "the appearance of a topic in a document stream is signaled by a 'burst of activity'" Kleinberg proposed probabilistic automata to formally model chronological sequences of documents as state machines [8]. He introduces a two-state and an infinite state hidden Markov model (HMM) in which time gaps between consecutive messages correspond to states. The importance of this "burstiness" characteristic for understanding the impact of individual human behavior on higher-order phenomena has seen significant adoption [9]. We suggest expanding this idea to incorporate more complex cascade properties to construct states. Beyond measuring bursts by the time delta between individual resources, bursts can also be measured by the delta between reoccurring unique network motifs. Additional inherent properties of individual resources such as the publishing host extracted from their URI or the language of the content could also be taken into account.

Ultimately, the study of socio-technical computation promises to offer new insight towards information theory across the Web. Hitherto unobtainable connections between disparate local efforts might be traced back to a real-world stimulus, even if these efforts used varying identifiers for the same meme or topic. A global view of, say, efforts to understand and respond to the financial crisis of 2008 might emerge from the information cascading from central events.

References

- [1] Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2004, May). Information diffusion through blogspace. In *Proceedings of the 13th international conference on World Wide Web* (pp. 491-501). ACM.
- [2] Adar, E. and Adamic, L. A. 2005. Tracking Information Epidemics in Blogspace. In *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI '05)*.
- [3] Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N. S., & Hurst, M. (2007, April). Patterns of Cascading behavior in large blog graphs. In *SDM* (Vol. 7, pp. 551-556).
- [4] Leskovec, J., Backstrom, L., Kleinberg, J. 2009. Meme-tracking and the Dynamics of the News Cycle. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD).
- [5] Goel, S., Watts, D. J., & Goldstein, D. G. (2012, June). The structure of online diffusion networks. In *Proceedings of the 13th ACM conference on electronic commerce* (pp. 623-638). ACM.
- [6] Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., & Leskovec, J. (2014, April). Can cascades be predicted?. In *Proceedings of the 23rd international conference on World Wide Web* (pp. 925-936).
- [7] Qu, Q., Liu, S., Jensen, C. S., Zhu, F., & Faloutsos, C. (2014). Interestingness-Driven Diffusion Process Summarization in Dynamic Networks. In *Machine Learning and Knowledge Discovery in Databases* (pp. 597-613). Springer Berlin Heidelberg.
- [8] Kleinberg, J. (2003). Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, 7(4), 373-397.
- [9] Barabasi, A. L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039), 207-211.