Patterns of Influence in a Recommendation Network*

Jure Leskovec[1], Ajit Singh[1], and Jon Kleinberg[2,**]

[1] School of Computer Science, Carnegie Mellon University
{jure, ajit}@cs.cmu.edu
[2] Department of Computer Science, Cornell University
kleinber@cs.cornell.edu

Abstract. Information cascades are phenomena in which individuals adopt a new action or idea due to influence by others. As such a process spreads through an underlying social network, it can result in widespread adoption overall. We consider information cascades in the context of recommendations, and in particular study the patterns of cascading recommendations that arise in large social networks. We investigate a large person-to-person recommendation network, consisting of four million people who made sixteen million recommendations on half a million products. Such a dataset allows us to pose a number of fundamental questions: What kinds of cascades arise frequently in real life? What features distinguish them? We enumerate and count cascade subgraphs on large directed graphs; as one component of this, we develop a novel efficient heuristic based on graph isomorphism testing that scales to large datasets. We discover novel patterns: the distribution of cascade sizes is approximately heavy-tailed; cascades tend to be shallow, but occasional large bursts of propagation can occur. The relative abundance of different cascade subgraphs suggests subtle properties of the underlying social network and recommendation process.

1 Introduction

The social network of interactions among a group of individuals plays a fundamental role in the spread of information, ideas, and influence. Such effects have been observed in many cases, when an idea or action gains sudden widespread popularity through word-of-mouth or "viral marketing" effects. To take a recent example from the technology domain, free e-mail services such as Microsoft's Hotmail and later Google's Gmail achieved wide usage largely through referrals, rather than direct advertising. (Gmail achieved wide usage at a time when the

* Work partially supported by the National Science Foundation under Grants No. IIS-0209107 IIS-0205224 INT-0318547 SENSOR-0329549 EF-0331657IIS-0326322 CCF-0325453, IIS-0329064, CNS-0403340, CCR-0122581, a David and Lucile Packard Foundation Fellowship, and also by the Pennsylvania Infrastructure Technology Alliance (PITA). This publication only reflects the authors' views.

** The work of the third author was performed in part while on sabbatical leave at Carnegie Mellon University.

W.K. Ng, M. Kitsuregawa, and J. Li (Eds.): PAKDD 2006, LNAI 3918, pp. 380–389, 2006.

only way to obtain an account was through a referral.) One also finds many examples in weblogs (blogs), where a piece of information spreads rapidly from one blogger to another before eventually being picked up by the mass media.

Information cascades are phenomena in which an action or idea becomes widely adopted due to influence by others [3, 5, 6]. Cascades are also known as "fads" or "resonance." Cascades have been studied for many years by sociologists concerned with the *diffusion of innovation* [15]; more recently, researchers in several fields have investigated cascades for the purpose of selecting trendsetters for viral marketing [9, 14], finding inoculation targets in epidemiology [13], and explaining trends in blogspace [2, 7, 10]. Despite much empirical work in the social sciences on datasets of moderate size, the difficulty in obtaining data has limited the extent of analysis on very large-scale, complete datasets representing cascades. Here we look at the patterns of influence in a large-scale, real recommendation network and examine the topological structure of cascades.

We address a set of related questions: What kinds of cascades arise frequently in real life? Are they like trees, stars, or something else? And how do they reflect properties of their underlying network environment? We describe (in Section 3) a large person-to-person recommendation network, consisting of 4 million people who made 16 million recommendations on half a million products. To analyze the data, we first create graphs where incoming edges influence the creation of outgoing edges. Then, we enumerate and count all possible cascade subgraphs, using an algorithm developed in Section 4. There, we propose an approximate heuristic for graph isomorphism involving the degree distribution and the eigenvalues of the adjacency matrix that scales to large datasets. We apply the algorithm to the recommendation dataset, and analyze it in Section 5.

We find novel patterns, and the analysis of the results gives us insight into the cascade formation process. We find that the distribution of cascade sizes can be approximated by a heavy-tailed distribution. Generally cascades are shallow but occasional large bursts also occur. The cascade sub-patterns reveal mostly small tree-like subgraphs; however we observe differences in connectivity, density, and the shape of cascades across product types. Indeed, the frequency of different cascade subgraphs is not a simple consequence of differences in size or density; rather, we find instances where denser subgraphs are more frequent than sparser ones, in a manner suggestive of properties in the underlying social network and recomendation process.

2 Related Work

To our knowledge, this is the first large-scale study of cascades in a real recommendation network. We believe the lack of prior studies is due in large part to the difficulty in acquiring large recommendation network datasets without link ambiguity from a real-world setting.

Most work on extracting cascades from large-scale on-line data has been done in the blog domain [1, 7, 10]. The authors in this domain note that, while information propagates between blogs, examples of genuine cascading behavior appeared