Part I → Part II → Part III → Part IV → Part V

Challenges

[Everyone's an influencer: quantifying influence on twitter]
bakshy, JM Hofman, [WA Mason], [DJ Watts]
WSDM 2011

People used to say half of the money spent in advertising is wasted, the problem is they do not know which half.
Measuring the return on investment of viral advertising is even more difficult, we have seen that there are many confounding factors. We only have anecdotes that most of the time describe exceptional cases. Thorough independent studies that compare the return on investment of viral campaigns versus traditional ones have not been carried out. Until those studies exist, we cannot assert that shifting spending from traditional marketing to viral marketing is a solid business proposition.

Prove Value for Social Sciences

- Online data may be huge but it is often **neither representative nor complete** (Ask a political scientist what she thinks about your election predictions with Twitter!)
- Offline data is difficult to obtain
  - External influence, e.g., mass media
- The main concern is the input data, how do we address it?

Surveys on predicting elections with Twitter.

Daniel Gayo-Avello: **"I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper" -- A Balanced Survey on Election Prediction using Twitter Data.** April 2012. http://arxiv.org/abs/1204.6441

Alexander Furnas: "**You Can't Use Twitter to Predict Election Results**". The Atlantic. May 2012. http://www.theatlantic.com/technology/archive/2012/05/you-cant-use-twitter-to-predict-election-results/257201/#

We now have access to enormous datasets containing the opinions of millions of people about political candidates, products, brands, etc. By now we should be able to predict everything, from presidents to product flops. But ...
The samples we obtain from social media have numerous biases, for instance, they are biased towards people who have adopted certain technologies. To put it bluntly, this usually means urban, educated, affluent, and of certain racial or cultural b/g.
Polling the opinions of people has always been a complex business, precisely to avoid introducing biases. Conclusions drawn from biased data samples, no matter how large they are, are of limited value.

Learn to Design for Virality

- What makes a product/idea/technology viral?
  - Role of content?
  - Role of seeds?
  - Other factors?
- For every video (post) that goes viral on YouTube (Twitter), hundreds fizzle out!
- How can we ***design*** a product or *meme* so that it is intrinsically "**sticky**"?
- Beyond anecdotes, **what do we know about the factors behind successful viral campaigns?**

We should be at least as interested in making things happen (e.g., go viral) as we have been in analyzing viral phenomena.
What makes a product/idea spread virally? How do you isolate the contributing factors?
What would be some appropriate ***data sets*** to learn the answers to the above questions?
How do put what we learn into the "market place"? How do we "manufacture" products/ideas so they go viral w.h.p.? Your product/idea will be ***competing*** with numerous others.

Algorithmic Challenges

- $O(|V|^2)$ algorithms considered not feasible for large graphs (e.g. $|V| > 1M$)
  - greedy IM algorithm, $\Omega(|V|^2)$
  - all-pair shortest paths or graph diameter, $\Omega(\frac{|V|^2}{\log|V|})$
  - betweenness centrality, $\Omega(|V|^2)$
- Need near-linear time algorithms
  - $O(|V| \text{ polylog}(|V|))$ algorithms
  - may need new algorithm paradigm (e.g. Laplacian paradigm [Chrstiano et al. STOC 2011, Spielman & Teng, SIAM JC 2011])
  - may need new complexity research on graph problems

Laplacian Paradigm:

Paul Christiano, Jonathan A. Kelner, Aleksander Madry, Daniel A. Spielman, Shang-Hua Teng: Electrical flows, laplacian systems, and faster approximation of maximum flow in undirected graphs. STOC 2011: 273-282

Daniel A. Spielman, Shang-Hua Teng: Spectral Sparsification of Graphs. SIAM J. Comput. 40(4): 981-1025 (2011).

More technical challenges

- Competitive diffusion
  - need more realistic model of competitive diffusion
    - validation by real-world traces
  - need incorporation of individual rationality
    - rationality of individuals in social networks
    - rationality of competing companies
- Adaptive viral marketing
  - use the effect of past diffusion or current partial diffusion to guide further seeding choice
- Handling dynamic changes in social networks
  - network structure, influence strength may change over time

There are more challenges at the technical level of studying information and influence diffusion and rival marketing appilcations

For example, for competitive diffusion, various competitive diffusion models still need to be validated by real-world data, and be further refined. It also needs to incorporate individual rationality, using methodologies from game theory and economics research. Some existing work has looked into some of the issues, e.g. Immorlica et al., N. Immorlica, J. Kleinberg, M. Mahdian, and T. Wexler. The role of compatibility in the diffusion of technologies in social networks, EC 2007, Goyal and Kearns, Competitive contagion in networks. STOC 2012.

Also, social networks are very dynamic, user interactions and relationships keep change. Thus future study of diffusion modeling, and influence maximization need to incorporate such network dynamics to reach more realistic results.

Another direction is looking into adaptive viral marketing, while the viral marketing decisions are adapted and guided from the partially known diffusion of the current campaign, or the information from past campaigns. A recent work by Golovin and Krause, Daniel Golovin, Andreas Krause: Adaptive Submodularity: A New Approach to Active Learning and Stochastic Optimization. COLT 2010, has looked into adaptive influence maximization in a general adaptive submodularity framework.

Sanjeev Goyal, Michael Kearns: Competitive contagion in networks. STOC 2012: 759-774
Questions include: a) how network structures affect the efficiency (Price of Anarchy) of the game; b) what about multiple players? c) what about multi-stage game in which two firms gradually spend their seed budgets? etc.

Push Technology out to Applications Beyond Viral Marketing

- Case studies of successful deployment of Influence/Information Propagation/ Maximization Technology in:

- Rumor/Innovation spreading modeling, detection, containment

- Trend detection and prediction

- Infection propagation detection and containment.

It’s one thing to analyze data sets to try to understand how infection/innovation/rumor/etc. spread. It’s quite another to deploy IP/IM technology in production systems in apps beside VM.

We are talking serious (perhaps commercial) apps here. Who would pay for IM/IP kind of service in the context of these other app. areas? What’s the case for that? Can we interest doctors and health scientists/officials in IM kind of technology? How would you convince them of the value proposition?

And to Conclude

- Great advances in theory, analysis, algorithms related to viral phenomena.
- But **engineering** of viral phenomena (in the context of any of the apps we have mentioned) has yet to be taken out of the lab!
- Thanks!