

Spillovers in Networks and User-Generated Content

Michael Kummer (ZEW)

Discussion

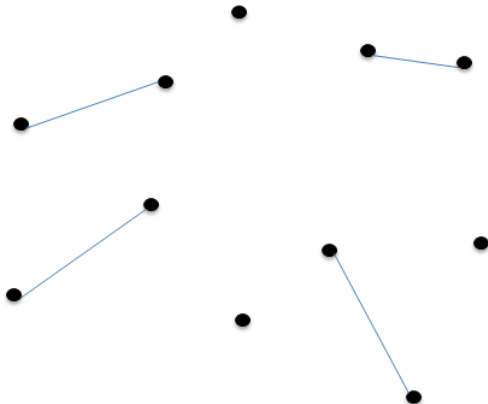
Pinar Yildirim (Wharton)

June 21, 2013

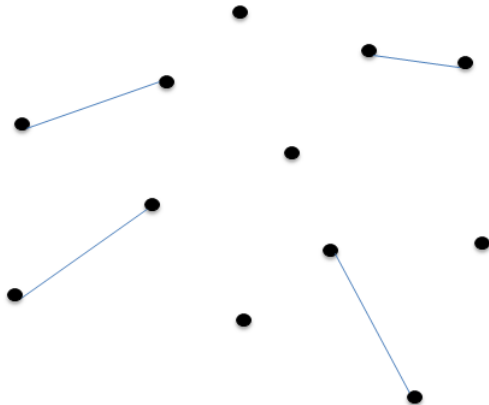
The Study & Background

- **User-generated Content (UGC)** and **formation of networks**
 - **Content & attention spillover** that results from *links* between network members (Mayzlin and Yoganarasimhan (2012), Goldenberg et al. (2010), Oestreicher-Singer and Sundararajan (2012))
 - Use **observable shocks** to analyze the spillover: natural disasters, featured articles, a product receiving a review
 - Using data from the **German Wikipedia**, considering revision of these links
- Ambitious, wants to go beyond the existing literature, & still under construction

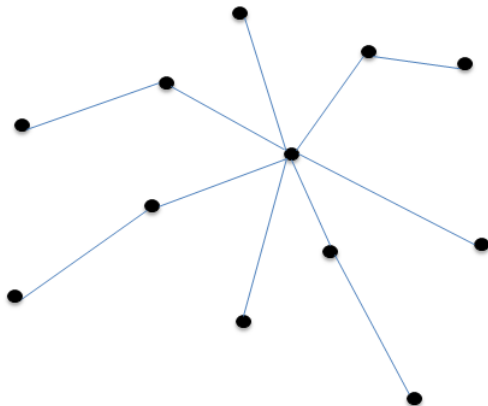
Setup



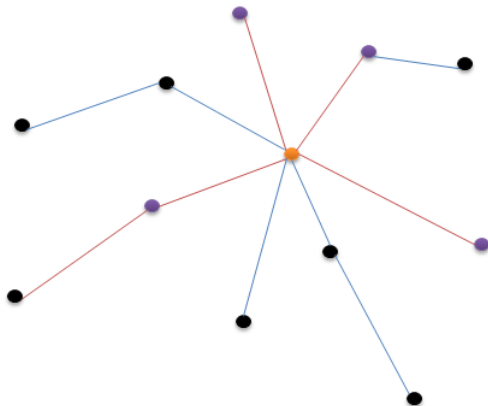
Setup

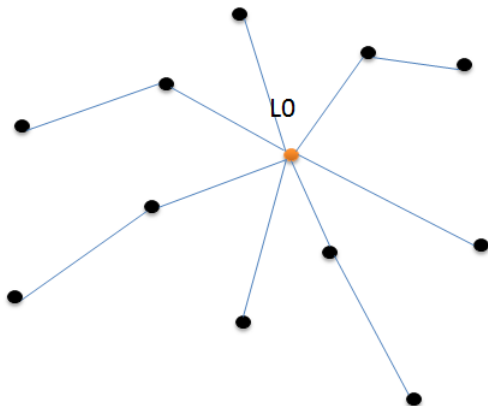


Setup

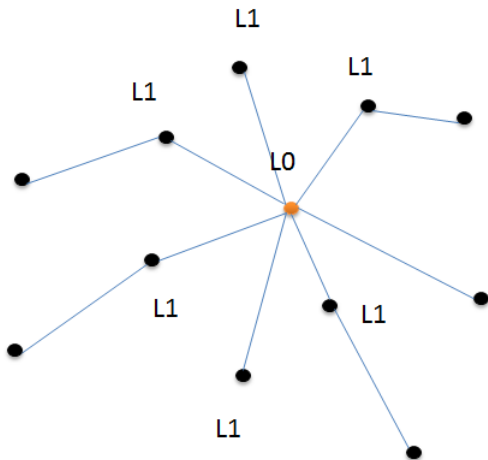


Setup

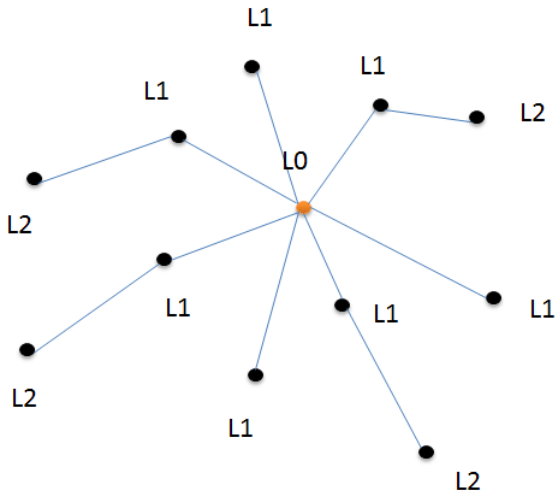




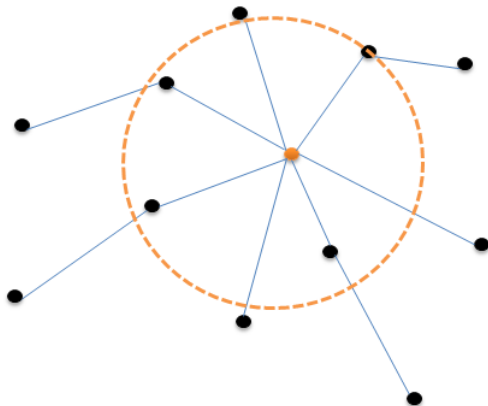
Spillover to immediate connections



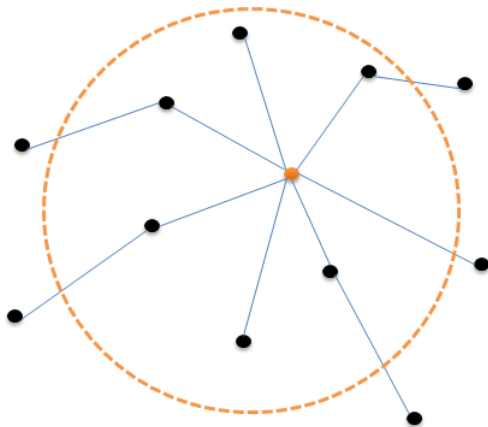
Spillover to 2nd degree connections



Limited Spillover of Attention



Larger Scale Spillover of Attention



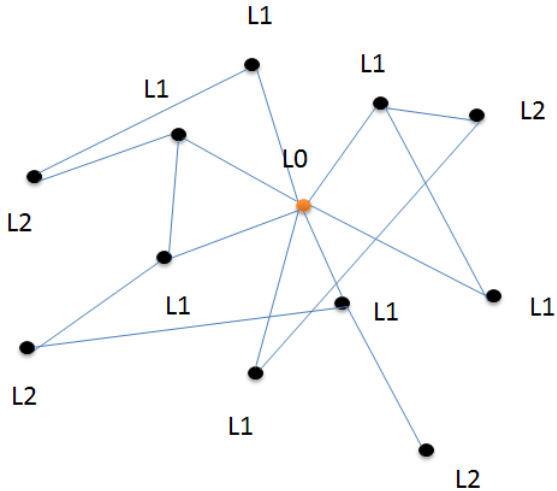
Links

Role of Links in Clicks and Content Update

- Use exogenous shocks to the network to investigate the role of links - one link suddenly receives 'attention' (natural disaster or an article being featured)
- Compare the increased interest with a 'control group'
 - Mixed evidence - both show increase in the clicks of neighboring nodes, weak evidence for content generation
- **Role of links** - Hard to suggest causality due to correlated demand for neighboring nodes

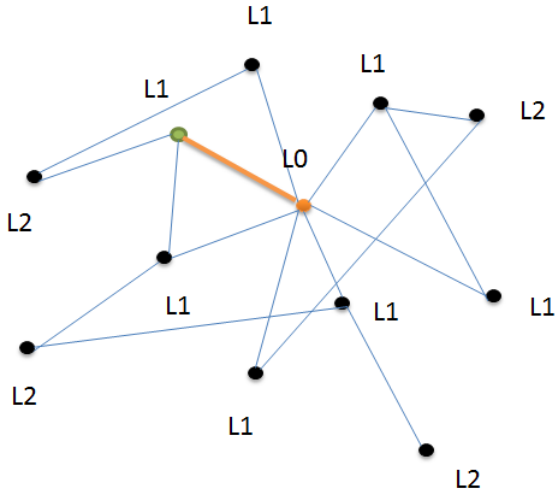
Network Structure

What happens when the L1 and L2 links are connected?



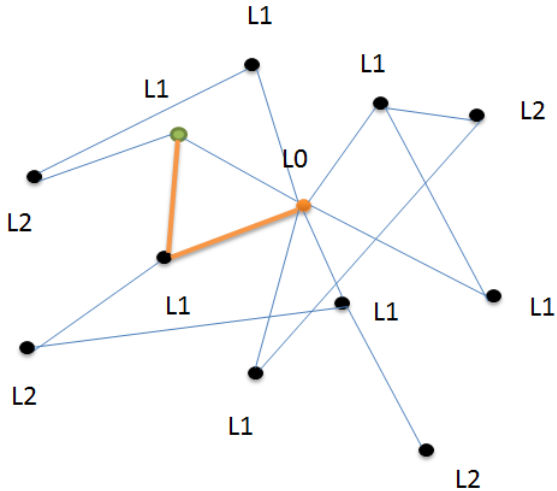
Network Structure

What happens when the L1 and L2 links are connected?



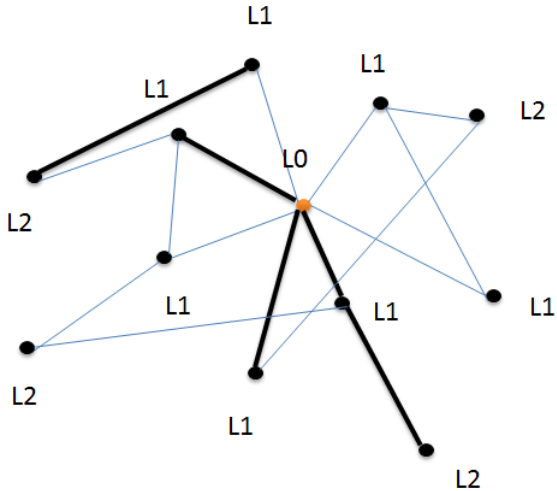
Network Structure

What happens when the L1 and L2 links are connected?



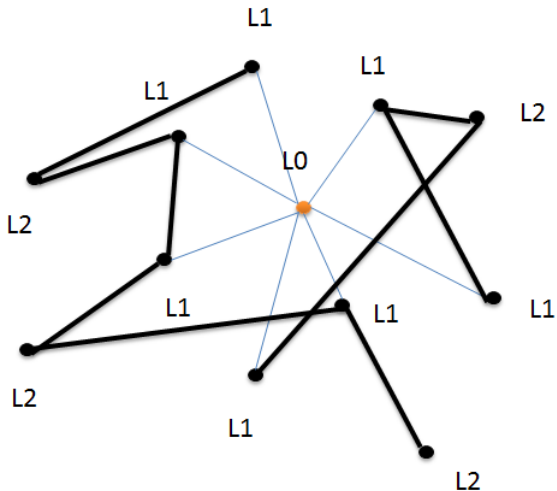
Network Structure

What happens when the edge characteristics are different?



Network Structure

...if not randomly different?



- Strong connection between network structure and the level of assortative mixing (link / relation characteristics)
 - **Distance**- number of links on the minimal path between two nodes
 - **Network Proximity** - consideration a weighted average of the lengths of **all** possible paths treated and neighboring node
 - **Local Clustering** - how close a node and its neighbors are to being a clique?
- Additional network structure should be added to the model

- There is a time order to the events - L1 and L2 links come before L0 links
 - ⇒ L0: Boxing Day Tsunami
 - ⇒ L1: Boxing Day
 - ⇒ L2: Indian Ocean
- **Direction** of spillover between the links - bidirectional edges?
- L1 - **very specific selection**, data ignores a part of the graph
 - Nodes that has a cite a week before the event happened

- Selection of the comparison dataset – **control group** is key
 - Tricky - 'pages similar but unlikely to be affected by the treatment' - **What is the similarity exactly?** Similarity in concepts but also network?
- Re-formation of the network post shock
 - If both G and X changes after the shock

Weak Results for Content Provision

- Weak evidence for content generation - is this an artifact of using data from Wikipedia?
- Content generation incrementally more **costly**
 - **Collective action**, some content will be substitutes, very closely **monitored**, under the control of a small group of moderators
- Reaching a global conclusion from **Wikipedia** [and natural disasters] about content generation is difficult. The impact of link structure may be more evident in other UGC sites.

- Data - very heavily skewed, long tailed
 - mean clicks: 35, max clicks: 20, 384
- X matrix includes a lot of important information not presented in the results
 - 'deleted links', 'links to further info'
- Error assumptions - remain similar overtime, assume pages remain identical after the shock

- Overall: ambitious & relevant
- When complete, the analysis may yield interesting results