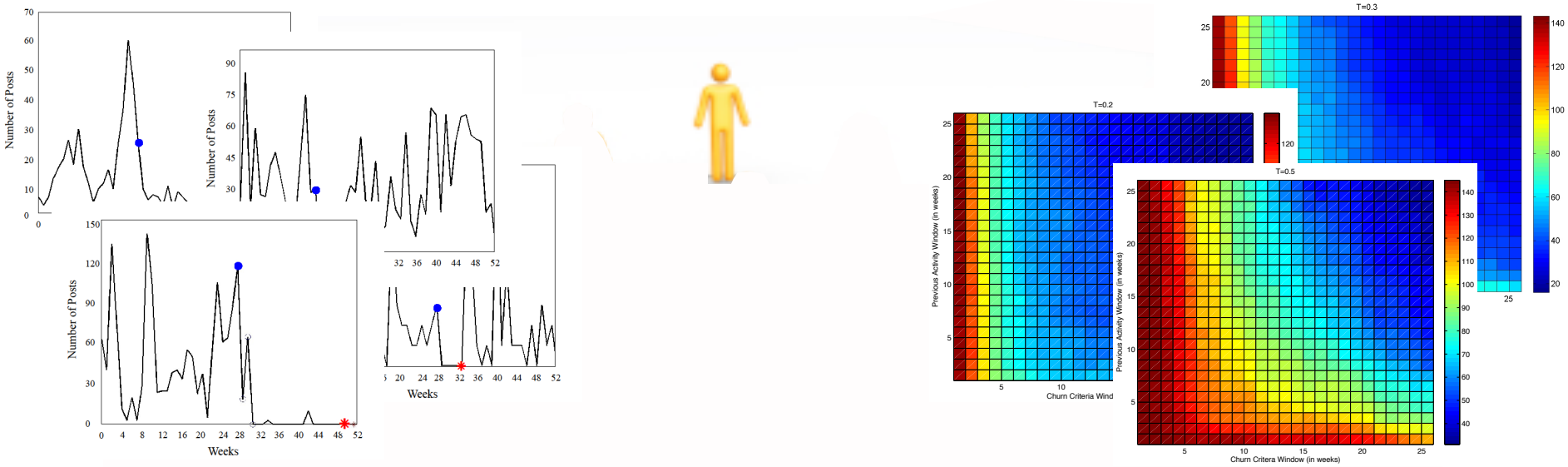
EC Project 257859

The Effect of User Features on Churn in Social Networks

Marcel Karnstedt, Matthew Rowe, Jeffrey Chan, Harith Alani and Conor Hayes

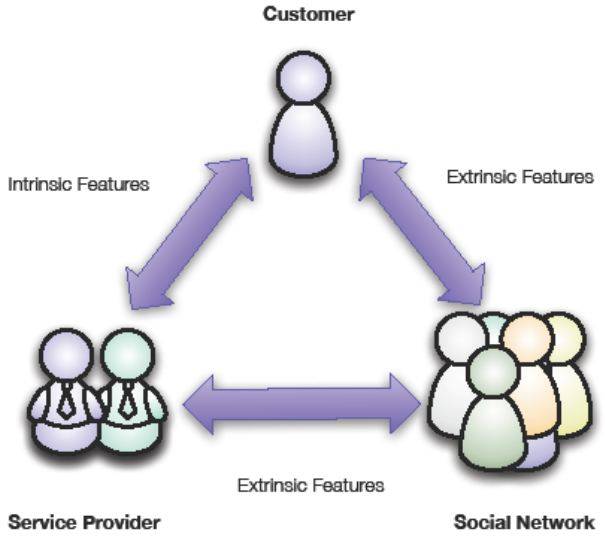
This work was partly funded by the European Union (EU) and Science Foundation Ireland (SFI) in the course of the projects ROBUST (EU grant no. 257859), WeGov (EU grant no. 248512), the CLIQUE Strategic Research Cluster (SFI grant no. 08/SRC/I1407), and LION-2 (SFI grant no. SFI/08/CE/I1380).

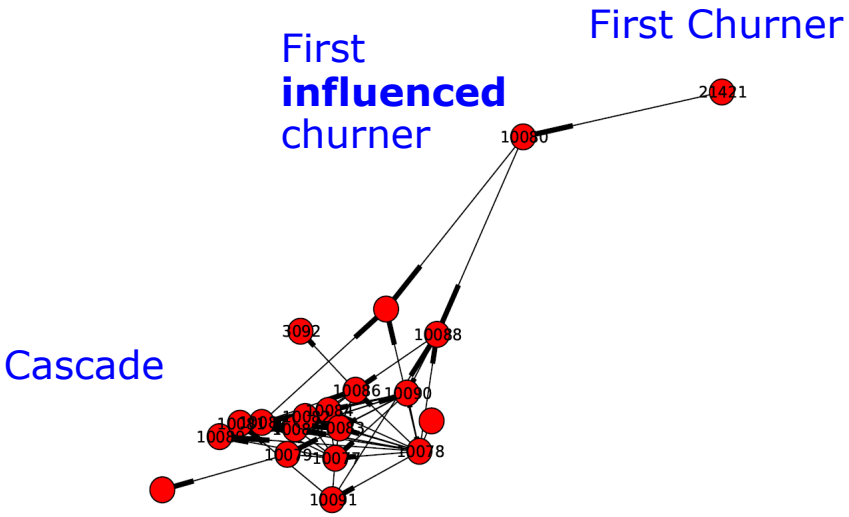
What is Churn??



- *Churn* is a risk (opposite to opportunities)
  - User retention vital to community health & functioning
- Hot topic in industries like telcoms
  - Not clearly defined in online social networks
  - Full defect vs. partial

Why Churn?





- Additional dimension
- Community & user value
- Personal needs & satisfaction
- Network effects observable

Where we are…

1. What is Churn?
   – Proposed flexible definition, different types
2. Why Churn?
   – Identified features, community & user value
3. **How to predict?!**
   – How do we know if a user is going to churn?
   – Can we correlate a user's value within a community with their churn probability?

- Set our future research agenda…

User Value & Churn Probability

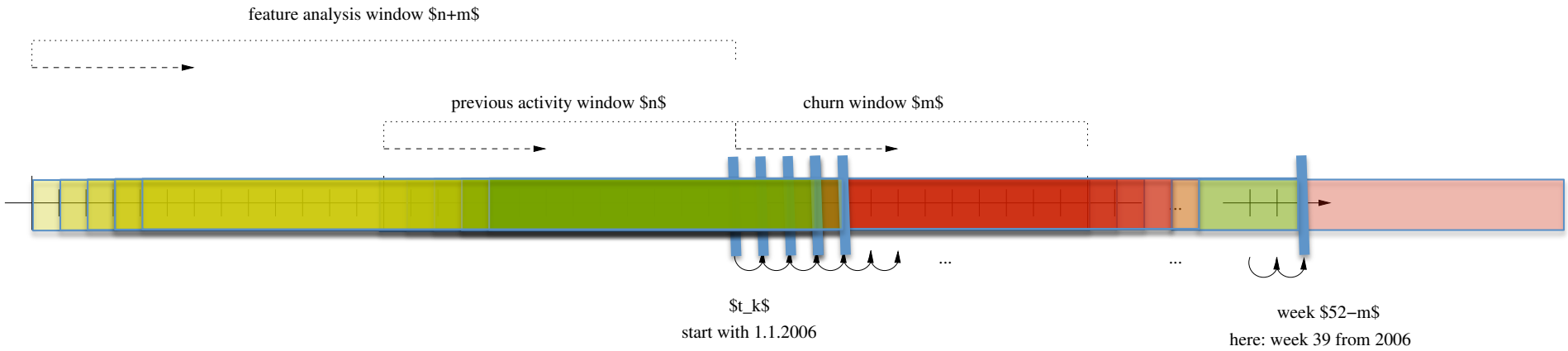
- Like in telcom: start with feature-based approach
- But we're sure: network effects have to be considered
- Correlation: features vs. churn probability

$$P(churn|v_i) = \begin{cases} 0 & \mu_C(v_i) \geq \mu_{PA}(v_i) \\ 1 - (\frac{\mu_C(v_i)}{\mu_{PA}(v_i)}) & otherwise \end{cases}$$

  - Features: structural and social network, reciprocity, persistence/productivity, popularity, sentiment, …

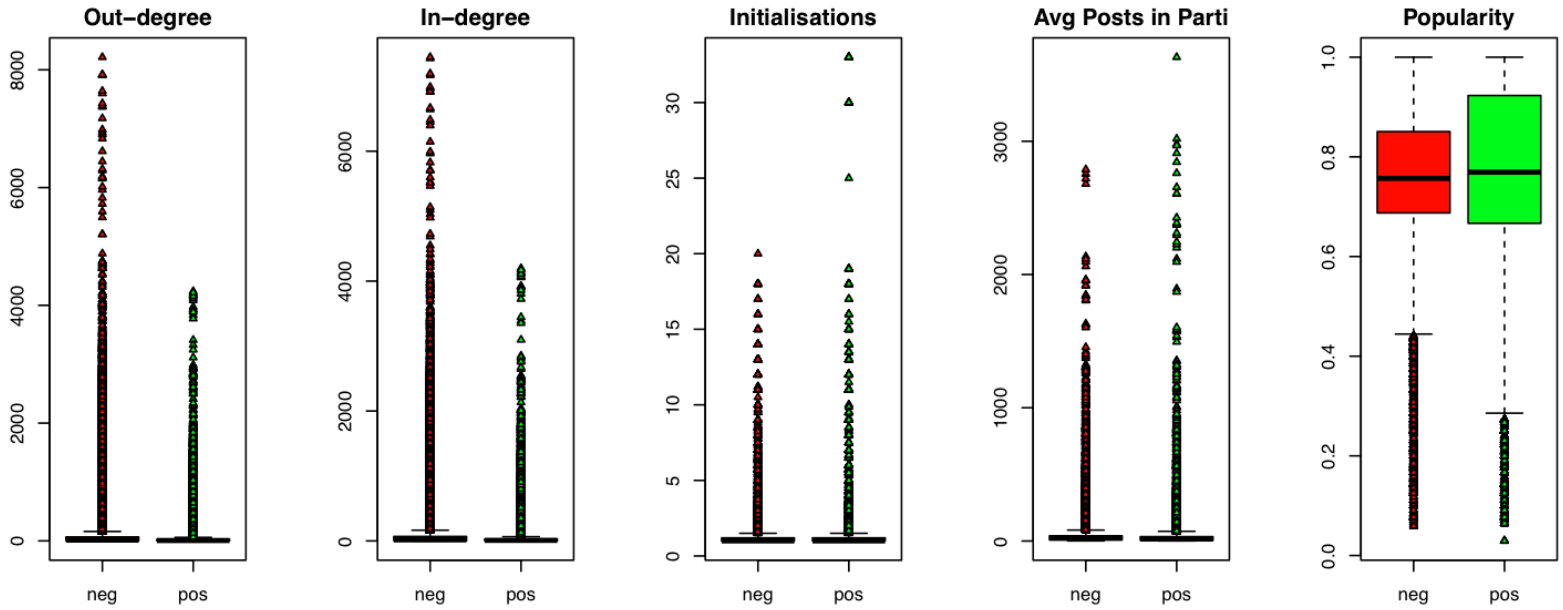
Case Study

- Data provided by the largest Irish message board Boards.ie
- Used all data published within 2006
- Derived features at weekly increments



Analysis: Global Churn

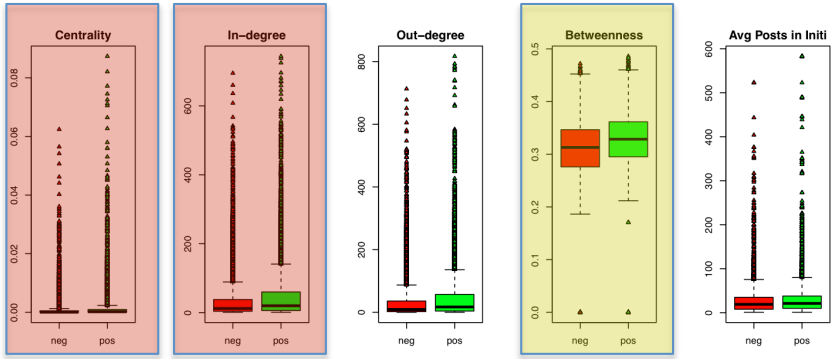
Threshold	P	R	F_1	κ
0.2	0.638	0.639	0.635	0.266
0.5	0.668	0.666	0.649	0.286
0.7	0.734	0.741	0.733	0.410



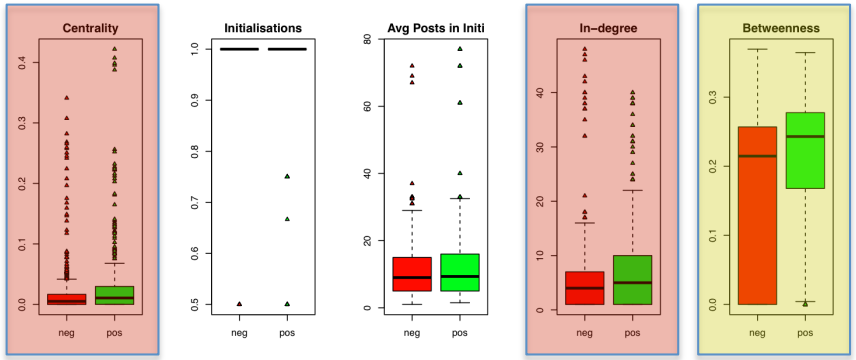
Expected!?

Surprise?!

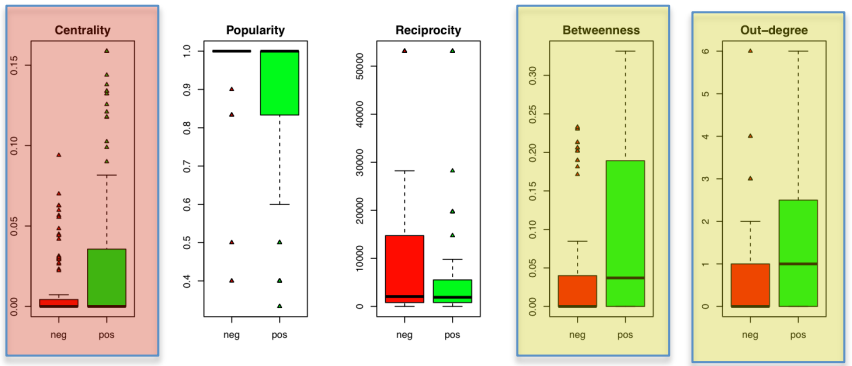
Analysis: Per Forum Churn



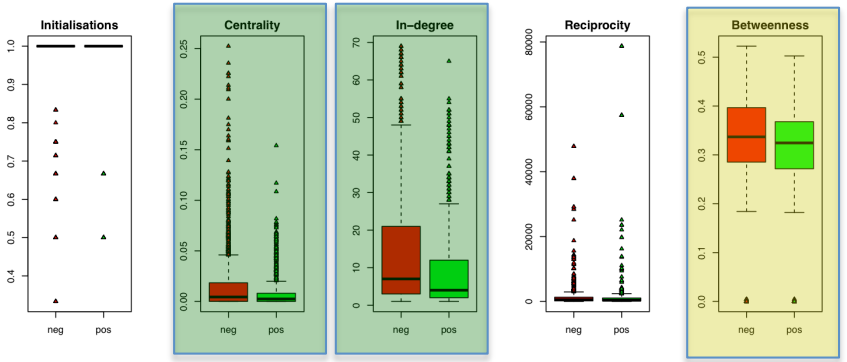
Highest: After hours



Median: Prime time cartoons



Lowest: Japanese culture



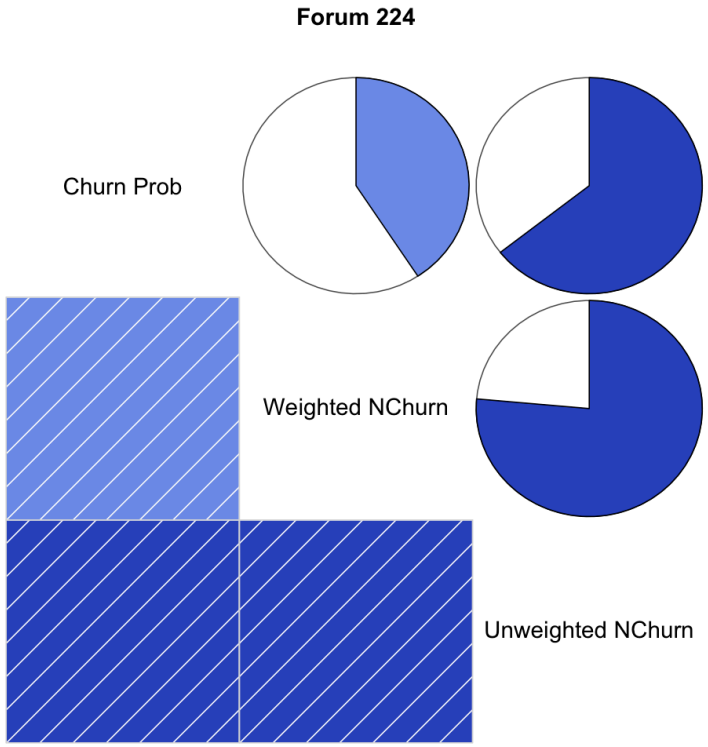
Mean: World of Warcraft

- Performance better than global, larger values better in forums with greater activity
- Distinct forums exhibit distinct behaviour

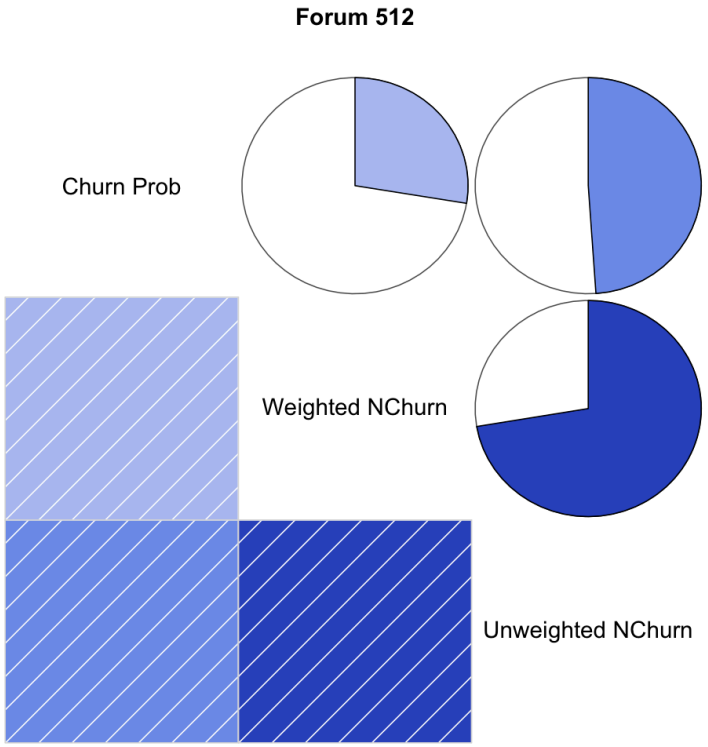
Towards Network Effects

- Unweighted Neighbourhood Churn:
  - Average churn probability of neighbours
- Weighted Neighbourhood Churn
  - Average weighted churn probability of neighbours
- Experiments over 4 forums from before
- Induced Linear Regression Model
  - Dependent variable: churn probability
  - Independent variables: unweighted and weighted neighbourhood churn probabilities

Analysis: Per Forum Neighbourhood Churn



Lowest: Japanese culture



Median: Prime time cartoons

Positive correlation between churn probability of a user and neighbourhood churns

…what about other features of the neighbourhood?!?

Conclusion

- Correlation existing, but significant differences:
  - Global vs. local
  - Between forums
- Advanced analysis of forum characteristics and these effects needed … classification/clustering
- Extend analysis (windows, filters, sample, …)
- Choice of features, integrate social roles
- Analysis of network effects
- *…value, health, personal needs, … loads to do!*