Measuring wiki viability

An empirical assessment of the social dynamics of a large sample of wiki-based communities

Dario Taraborelli

d.taraborelli@surrey.ac.uk

CRESS Seminar
November 4, 2008

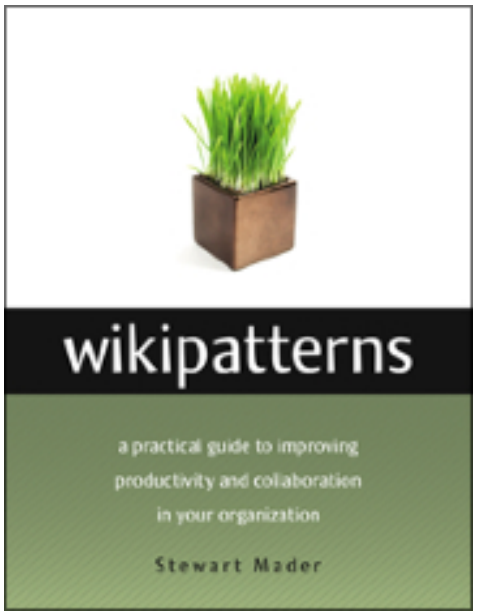
What factors determine the viability of online communities?

- Content-based online communities face a variety of **risks**:
  - content explosion
  - population extinction
  - low user turnover/high rate of user dropout
  - insufficient/unmanageable rate of user activity
- Typical solution: **best practices**

What factors determine the viability of online communities?

- Content-based online communities face a variety of **risks**:
  - content explosion
  - population extinction
  - low user turnover/high rate of user dropout
  - insufficient/unmanageable rate of user activity
- Typical solution: **best practices**

What factors determine the viability of online communities? (2)



"Wikipatterns is a toolbox of *patterns* and *anti-patterns*. Looking to spur wiki adoption? Want to grow from 10 users to 100, or 1,000? Applying patterns that help guide the growth of content, and recognizing anti-patterns that might hinder growth, can give your wiki the greatest chance of success"

What factors determine the viability of online communities? (3)

- **Wikipedia-centered literature** on wiki dynamics
- Scarcity of **data** on the temporal dynamics of such communities
- Lack of tools to make **empirically grounded predictions** on what determines the extinction or survival of an online community
- Little guidance for policy makers on the **governance** of such communities

What factors determine the viability of online communities? (3)

- **Wikipedia-centered literature** on wiki dynamics
- Scarcity of **data** on the temporal dynamics of such communities
- Lack of tools to make **empirically grounded predictions** on what determines the extinction or survival of an online community
- Little guidance for policy makers on the **governance** of such communities

What factors determine the viability of online communities? (3)

- **Wikipedia-centered literature** on wiki dynamics
- Scarcity of **data** on the temporal dynamics of such communities
- Lack of tools to make **empirically grounded predictions** on what determines the extinction or survival of an online community
- Little guidance for policy makers on the **governance** of such communities

What factors determine the viability of online communities? (3)

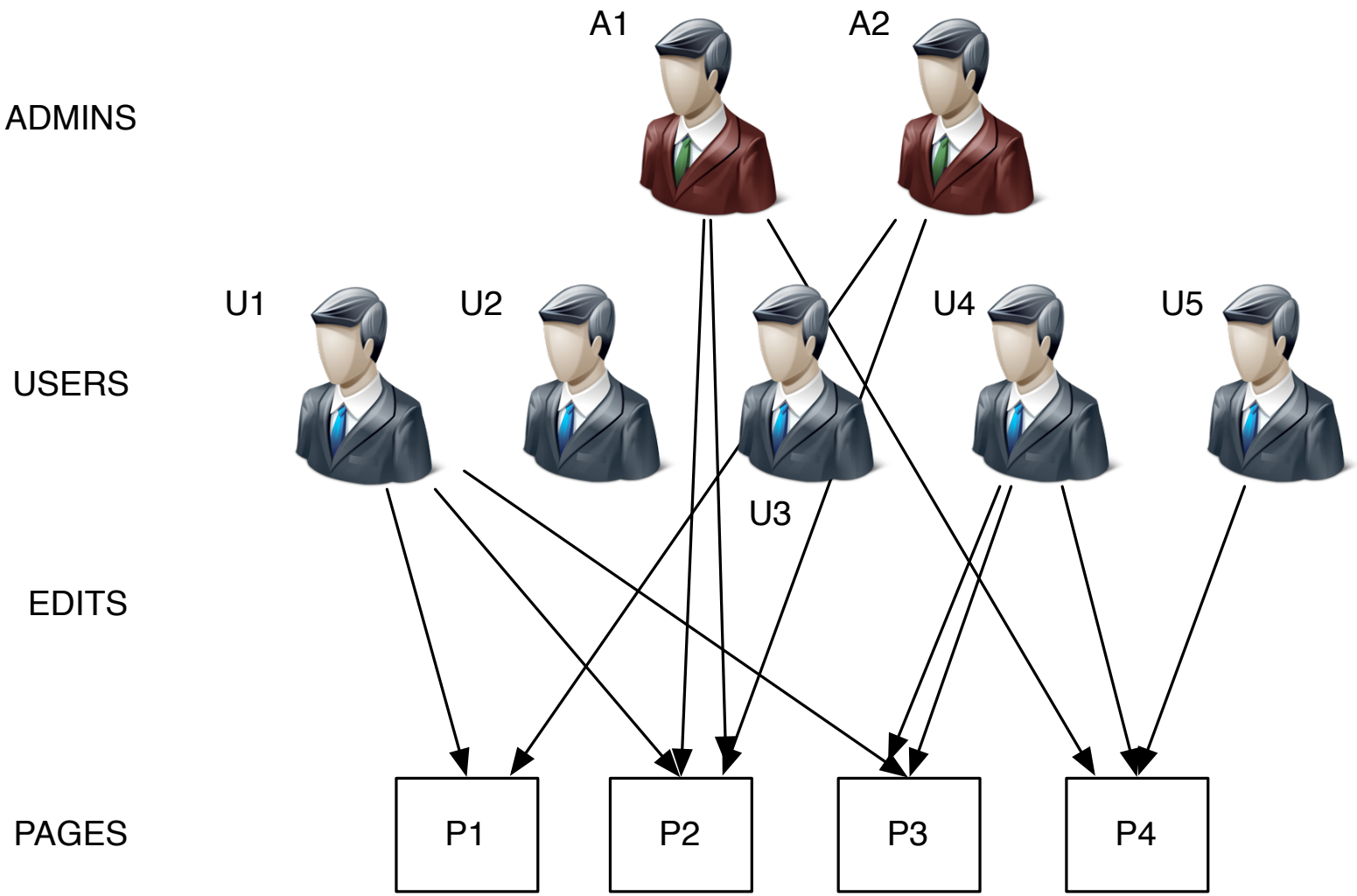
- **Wikipedia-centered literature** on wiki dynamics
- Scarcity of **data** on the temporal dynamics of such communities
- Lack of tools to make **empirically grounded predictions** on what determines the extinction or survival of an online community
- Little guidance for policy makers on the **governance** of such communities

Overview

Goal: Identify factors that affect the dynamics of a large sample of wiki-based communities.

1. Wiki dynamics: methodology
2. Results: growth enhancers and regulators
3. Research directions and conclusions

Method: Wiki activity scheme



Method: dataset

- **Data source**

  Growth data tracked over an 8-month span from $11,500+$ MediaWikis

- **Selected sample**

  360 wikis, with an initial population between 400 and $20,000$ users, restricted to hosters with reliable data, and with no major discontinuity in daily change rates.

- **Variables**
  - population size (U)
  - content size (P)
  - admin population (A)
  - edits (E)
  - access control (R)

Method: dataset

- **Data source**

  Growth data tracked over an 8-month span from $11,500+$ MediaWikis

- **Selected sample**

  360 wikis, with an initial population between 400 and $20,000$ users, restricted to hosters with reliable data, and with no major discontinuity in daily change rates.

- **Variables**
  - population size (U)
  - content size (P)
  - admin population (A)
  - edits (E)
  - access control (R)

Method: dataset

- **Data source**

  Growth data tracked over an 8-month span from $11,500+$ MediaWikis

- **Selected sample**

  360 wikis, with an initial population between 400 and $20,000$ users, restricted to hosters with reliable data, and with no major discontinuity in daily change rates.

- **Variables**
  - population size (U)
  - content size (P)
  - admin population (A)
  - edits (E)
  - access control (R)

Method: Indicators considered

descriptive indicators

- **user activity**, i.e. the proportion of edits per user (E/U)
- **user density**, i.e. the proportion of users per page (U/P)
- **edit density**, i.e. the proportion of edits per page (E/P).

governance factors

- **administrator ratio**, i.e. the proportion of users with admin status (A/U)
- **administrator density**, i.e. the proportion of admins per page (A/P)
- **editing permission** (R).

Method: Indicators considered

descriptive indicators

- **user activity**, i.e. the proportion of edits per user (E/U)
- **user density**, i.e. the proportion of users per page (U/P)
- **edit density**, i.e. the proportion of edits per page (E/P).

governance factors

- **administrator ratio**, i.e. the proportion of users with admin status (A/U)
- **administrator density**, i.e. the proportion of admins per page (A/P)
- **editing permission** (R).

Method: Indicators considered

descriptive indicators

- **user activity**, i.e. the proportion of edits per user (E/U)
- **user density**, i.e. the proportion of users per page (U/P)
- **edit density**, i.e. the proportion of edits per page (E/P).

governance factors

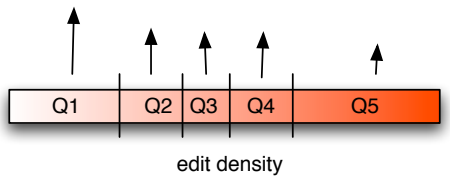
- **administrator ratio**, i.e. the proportion of users with admin status (A/U)
- **administrator density**, i.e. the proportion of admins per page (A/P)
- **editing permission** (R).

Method: Analysis

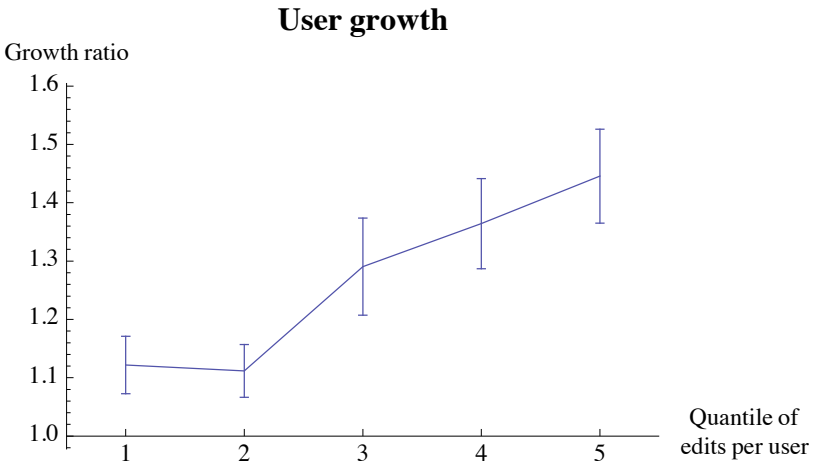
1. Sort wikis according to each of these indicators (e.g. *edit density*)

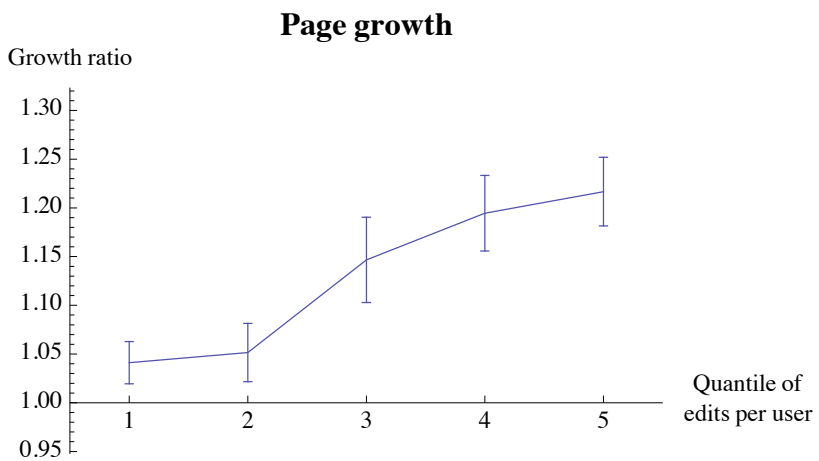
2. Aggregate wikis in bins (quantiles) of identical size (*e.g. wikis with low edit density vs. wikis with high edit density*)

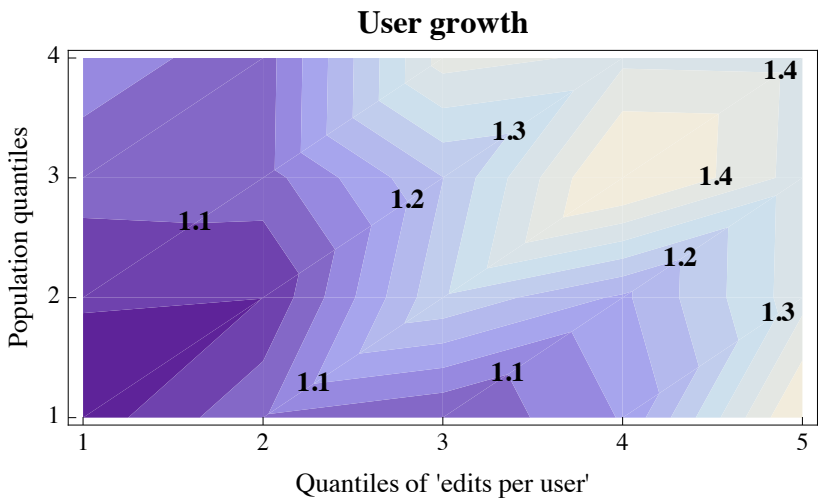
3. Measure how each quantile performs with respect to content and population growth rates

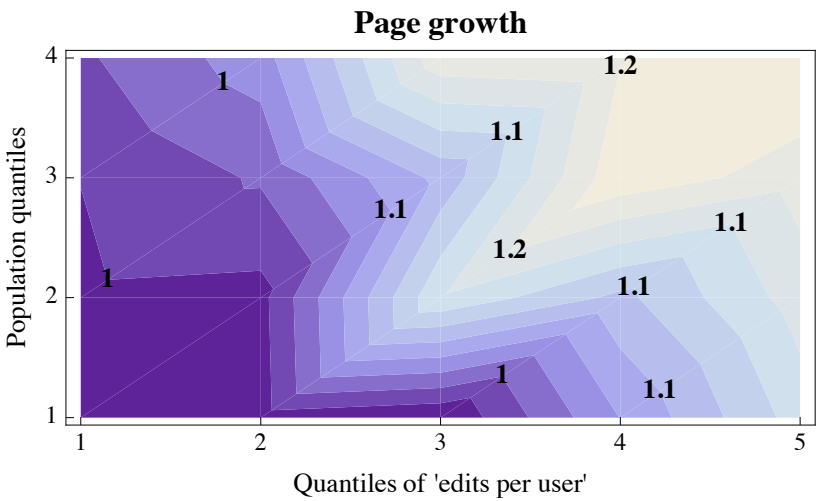


Growth enhancers (1): user activity

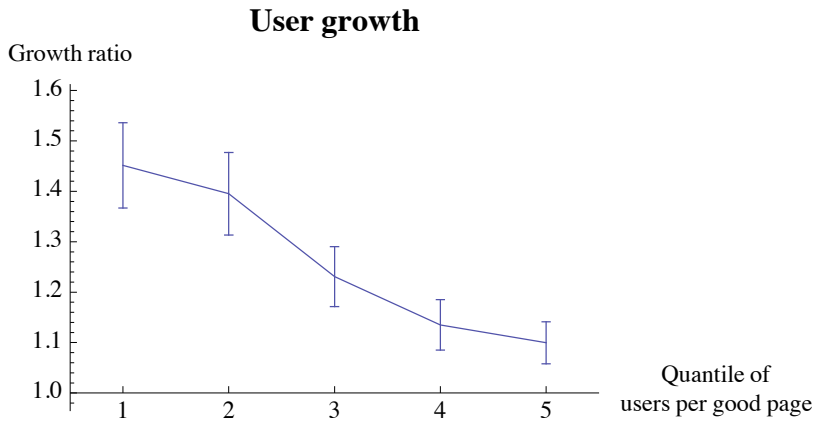


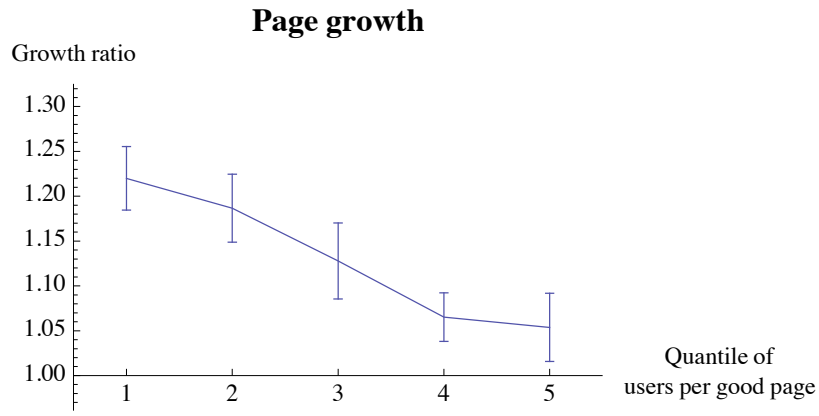


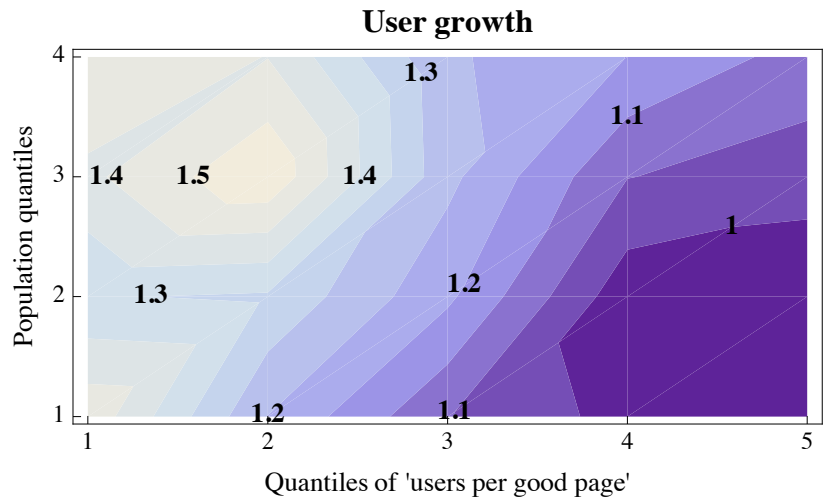


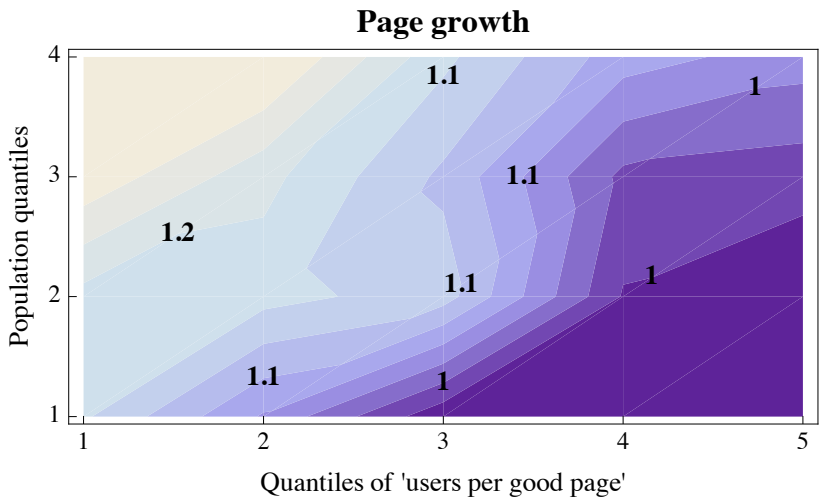


Growth regulators (1): user density

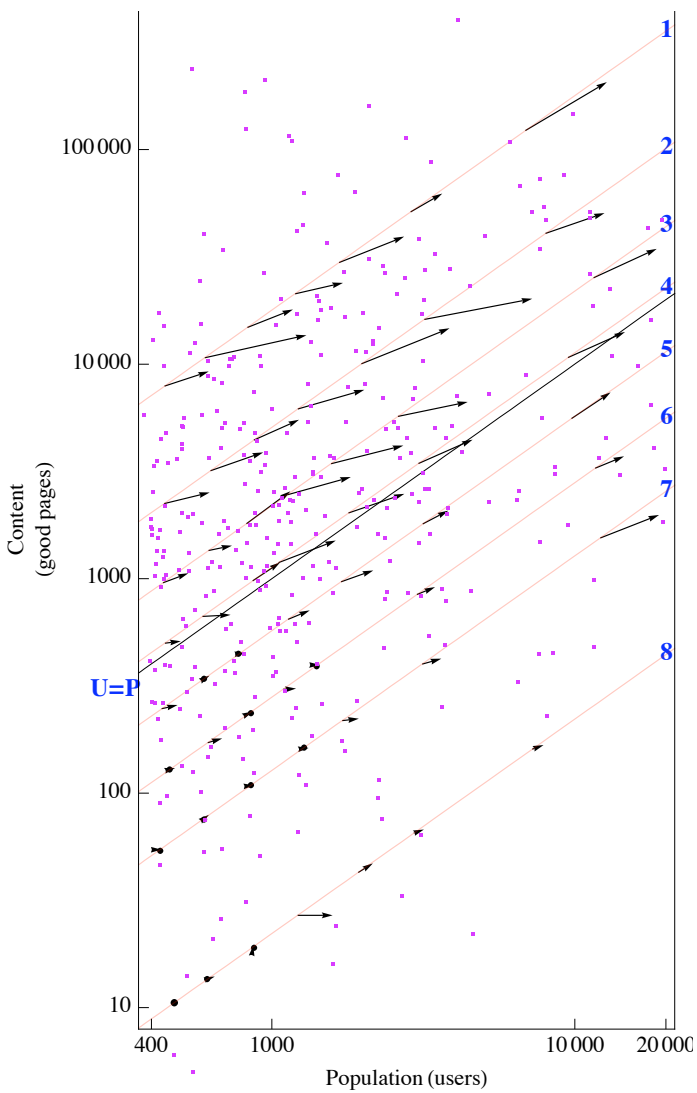


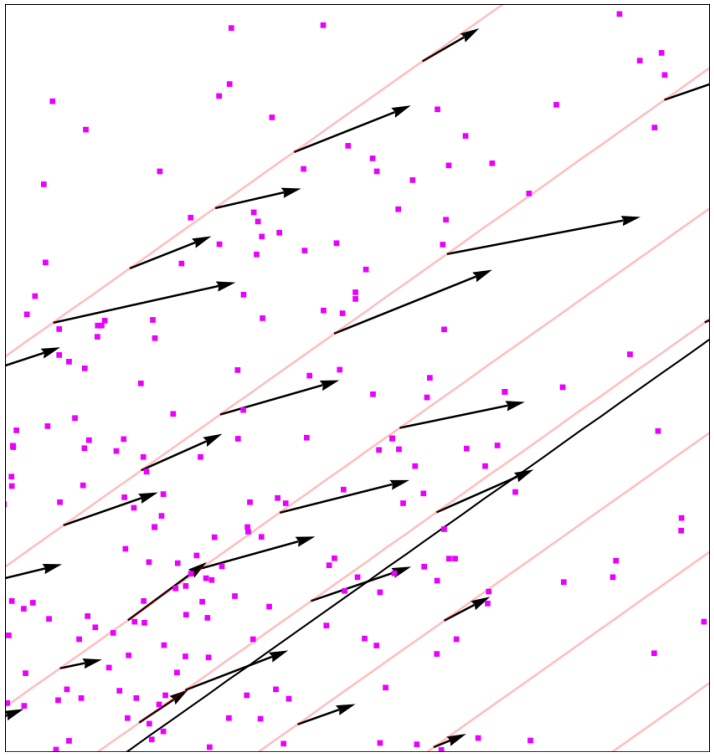






Wiki growth landscape (user density)





Growth enhancers (2): editing permission

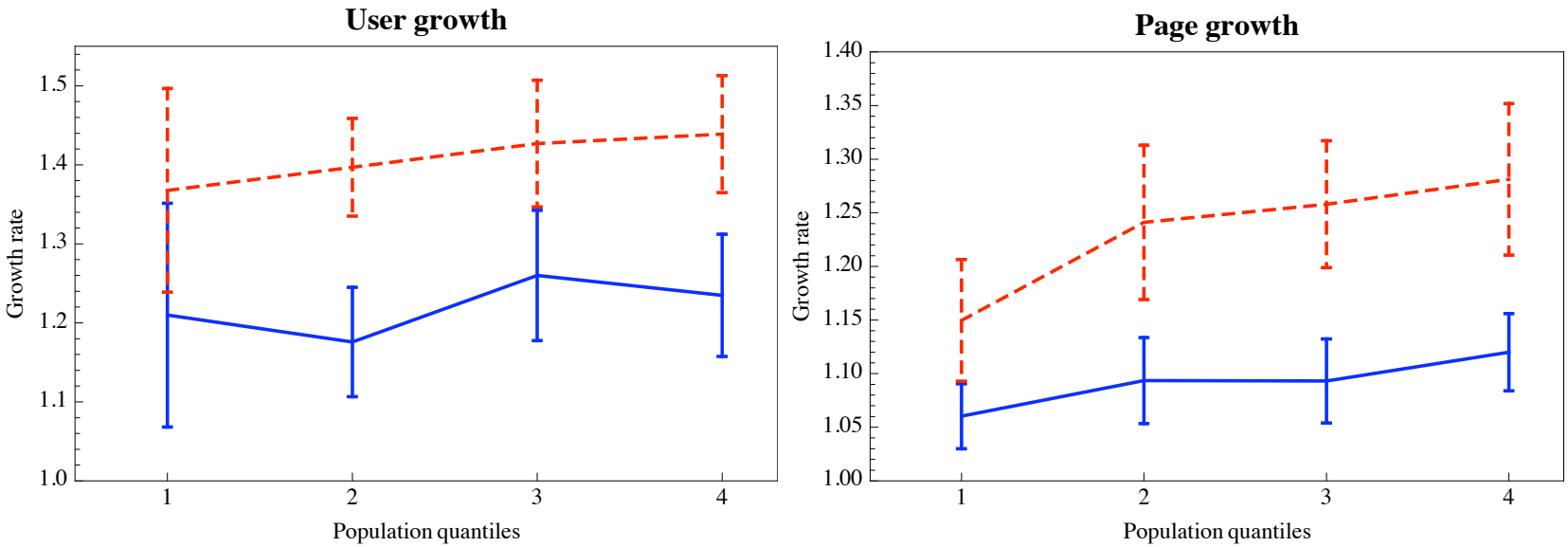
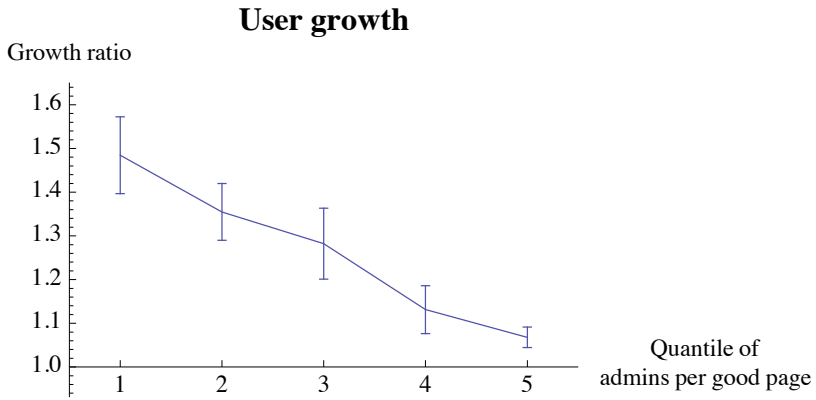
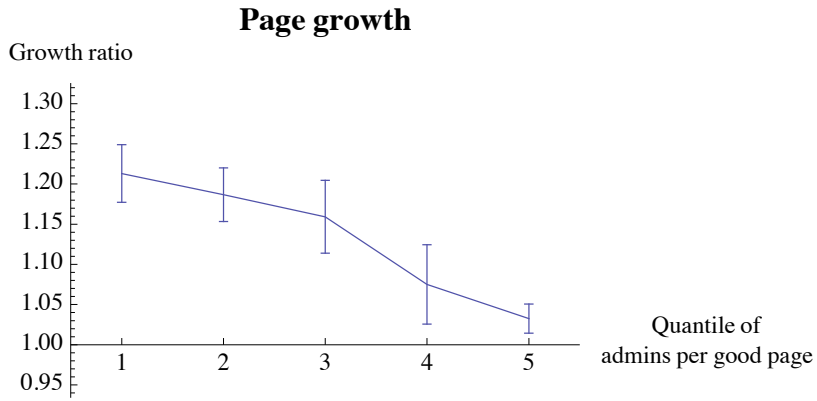
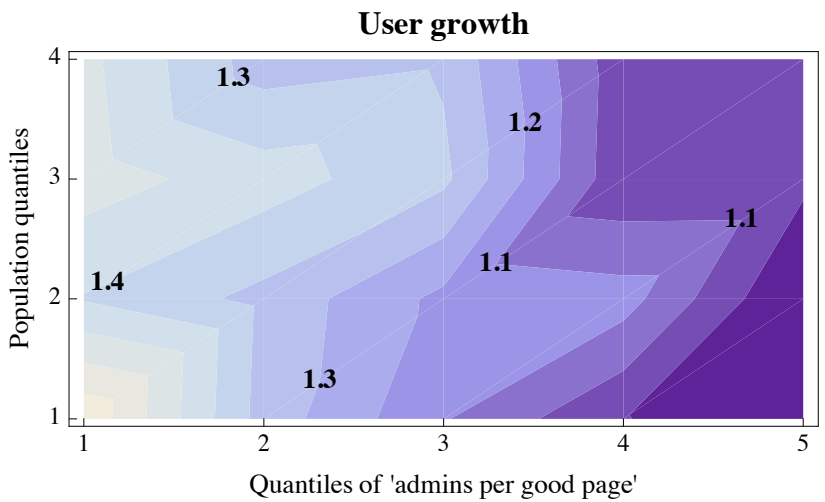


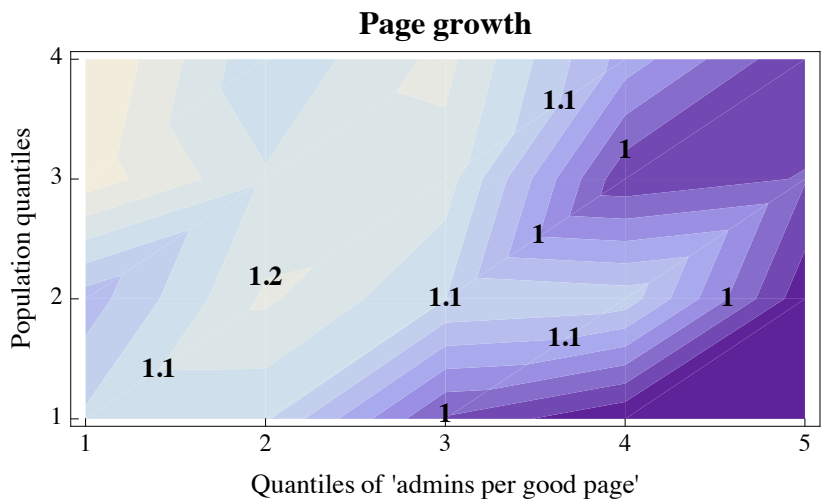
Figure 5: Growth landscape with respect to *editing permission*: *red dashed* refers to anonymously editable wikis, while *blue solid* to wikis editable by registered users only.

Growth regulators (2): admin density

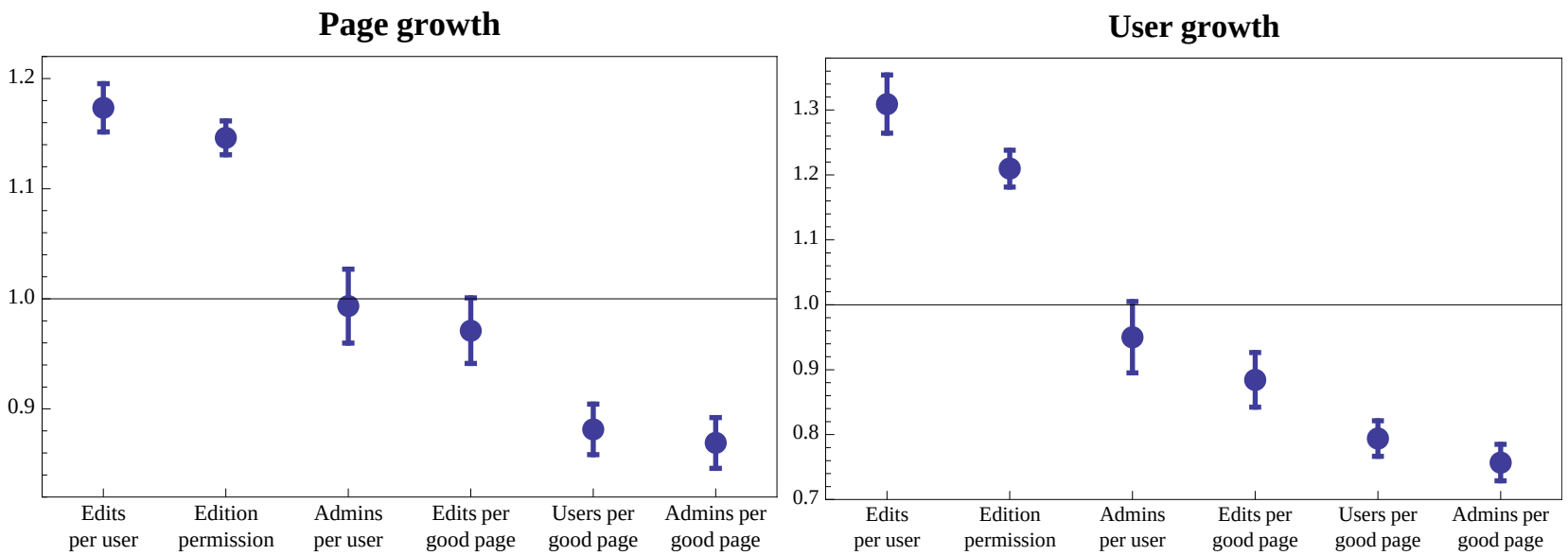








Summary of results



Summary of results

	Variable	*Growth rate*	
		Population	Content
STRUCTURAL INDICATORS	User activity (E/U)	+ +	+ +
	Edit density (E/P)	-	—
	User density (U/P)	- -	- -
GOVERNANCE FACTORS	Editing permission (R)	+ +	+ +
	Admin ratio (A/U)	—	—
	Admin density (A/P)	- -	- -

Table 2: Effect of different factors on wiki growth rates.

Summary of results

structural indicators

- The higher the ratio of **edits per user** the faster the wiki grows. Wikis with very active user communities are more likely to attract a large number of new contributors.
- Opposite effect of high **user density** per page.

governance indicators

- Population growth is more than 20% faster for **anonymously editable** wikis. Less barriers favor population growth.
- Too many **administrators per page** may hinder the content growth of a wiki.
- The ratio of **administrators per user** does not show a significant influence on growth dynamics.

Summary of results

structural indicators

- The higher the ratio of **edits per user** the faster the wiki grows. Wikis with very active user communities are more likely to attract a large number of new contributors.
- Opposite effect of high **user density** per page.

governance indicators

- Population growth is more than 20% faster for **anonymously editable** wikis. Less barriers favor population growth.
- Too many **administrators per page** may hinder the content growth of a wiki.
- The ratio of **administrators per user** does not show a significant influence on growth dynamics.

Wiki viability: Beyond raw growth indicators

- Better metrics to assess the **viability** of a Web community
  - content persistence
  - vandalism detection time
  - evolution of page/stub ratio
  - page creation rate
- Need of more fine-grained **independent variables**:
  - rate, magnitude and temporal properties of disruptions
  - ratio of lurkers per active users
  - metacontent production indicators (talk pages, comments)
  - wiki structure (categories, namespaces)
  - social network indicators

Wiki viability: Beyond raw growth indicators

- Better metrics to assess the **viability** of a Web community
  - content persistence
  - vandalism detection time
  - evolution of page/stub ratio
  - page creation rate
- Need of more fine-grained **independent variables**:
  - rate, magnitude and temporal properties of disruptions
  - ratio of lurkers per active users
  - metacontent production indicators (talk pages, comments)
  - wiki structure (categories, namespaces)
  - social network indicators

Towards model-based predictions of wiki dynamics (1)

1. Collect large-scale data on wiki dynamics
2. Identify significant growth patterns
3. Build models of wiki dynamics and resilience against disruptions
4. Provide model-based predictions and governance recommendations

Towards model-based predictions of wiki dynamics (2)

WikiTracer: Mapping the wikisphere

- Web service providing *platform-independent analytics* and *comparative growth statistics* for wikis.
- Inspired by Flickr Group Trackr: tracking and analyzing the dynamics of 11,000 + public Flickr Groups since Jan 2007.
- Idea first introduced at WikiSym '08

- Large support for major wiki engines:

MEDIAWIKI DOKUWIKI TWIKI XWIKI TIKIWIKI
WIKKAWIKI WACKOWIKI ODDMUSE

The end

Funding

- PATRES: Pattern Resilience (NEST-043268)

References

- Roth, C., Taraborelli, D., and Gilbert, N. (2008) Measuring wiki viability. An empirical assessment of the social dynamics of a large sample of wikis. *Proceedings of the 4th International Symposium on Wikis* - WikiSym 2008, New York, NY, USA. ACM.
- Roth, C., Taraborelli, D., and Gilbert, N. (2008) Démographie des communautés en ligne: le cas des wikis. *Réseaux* (forthcoming)
- Roth, C. (2007). Viable wikis: Struggle for life in the wikisphere. *Proceedings of the 3rd international symposium on Wikis* - WikiSym 2007, New York, NY, USA. ACM.