

Analysing and modelling editors' behavior of different languages in Wikimedia projects

Anita Chandra, Department of Computer Science and Engineering, IIT Patna
Email: anita.pcs15@iitp.ac.in

I. INTRODUCTION AND MOTIVATION

In this research work, we compare behaviors of editors in different languages in Wikimedia projects. We consider 12 Wiki datasets which include 4 different projects of Wikimedia namely Wikipedia, Wikibook, Wikinews, and Wiktionary in different languages. Here, we report similarities and differences in editing behaviors of english, french and german editors for Wikibook and Wiktionary only. Because of the multilingual nature of Wikis, millions of editors contribute from all across the continents who have diverse cultures and backgrounds. Consequently, this provides a scope of comparison among editors of different languages which also sheds light on multilingual and multicultural editors' behavior. Here, we attempt to show that the differences observed in the physical world, also exist in the virtual world.

II. PRELIMINARY RESULTS

Wiki data are represented as bipartite network, where two disjoint sets are editors and edited objects (articles/books/news/words) and links exist when an editor edits an article (excluding bot edits). We have given a growth model of evolution of Wiki networks. Since our interest lies in editors' behavior, we have therefore given the evolution of degree distribution of articles.

Model description: The wiki network is represented as triplet $G = \langle E, A, E_a \rangle$, where E and A are two disjoint sets of editors and articles respectively and $E_a \subseteq E \times A$ is set of edits to articles. Initially, we have e_0 editors in E , a_0 articles in A and e_{a0} edits in E_a . Let an article $a \in A$ such that k_a is degree of an article a . Network evolves as, at each time step, one editor and 'w' articles are added and edges arrive only from the editors set. Here, we assume that the arrival of next new editor specifies the next time step. Thus, in subsequent time step, a new editor becomes an old one. The parameters m and n denote the number of external edges brought by a new editor and number of internal edges from all old editors respectively. These edges are attached to articles with a combination of preferential and random mechanism. We have introduced two different randomness parameters γ_e and γ_i for new and old editors respectively. Attachment kernel for an edge is given as,

Attachment kernel: Let $\tilde{A}(k_{a,t})$ denotes the attachment probability of an incoming edge to an article a at time t where the degree of that article is $k_{a,t}$. Then $\tilde{A}(k_{a,t})$ is the attachment kernel and given as:

$$\tilde{A}(k_{a,t}) = \frac{k_{a,t} + \gamma}{\sum_{a=1}^{a_0+wt} (k_{a,t} + \gamma)}$$

From the description of model, change in degree of an article a is modelled as,

$$\frac{dk_a}{dt} = m \frac{k_a + \gamma_e}{\sum_{a=1}^{a_0+wt} (k_a + \gamma_e)} + n \frac{k_a + \gamma_i}{\sum_{a=1}^{a_0+wt} (k_a + \gamma_i)}$$

After Solving previous ordinary differential Eq., at initial condition $k_a(t_0) = 0$. We get,

$$p(k) = \frac{r}{s} (k_0)^{\frac{r}{s}} (k + k_0)^{-(1+\frac{r}{s})}$$

Where, $r = (c + w\gamma_e)(c + w\gamma_i)$, $s = mc + mw\gamma_i + nc + nw\gamma_e$, $k_0 = (m\gamma_e c + mw\gamma_e \gamma_i + n\gamma_i c + nw\gamma_e \gamma_i)/s$, $c = m + n$.

The results given in Fig 1 show good agreement between data and model with fitted parameters (estimated all the parameters of model i.e, m , n , w , γ_e and γ_i from datasets) for Wikibook (en,fr) and Wiktionary (en,fr,de). From the Fig 1a and 1b, we observe that $\gamma_i < \gamma_e$ in Wikibook (E) which means old English editors edit popular or well known books than new editors while in Wikibook (F) editing behaviors of french editors are opposite. From Fig 1c, 1d and 1e we have $\gamma_i > \gamma_e$ in every language of Wiktionary, that means new english, french and german editors always prefer to edit popular dictionary words while old editors edit less known words. Hence we observe different editing behaviors in new and old English and french editors of Wikibook while their patterns are similar in case of Wiktionary.

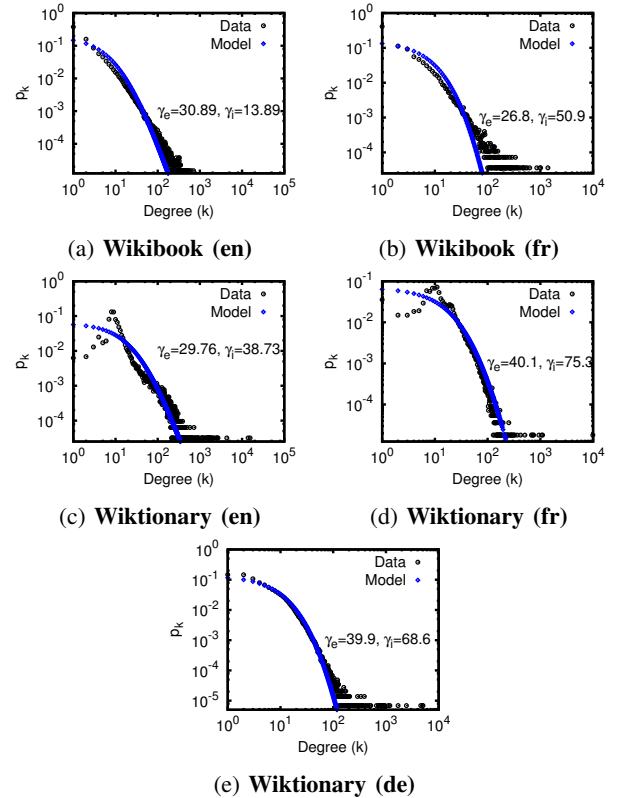


Fig. 1: (a),(b)- Books degree distributions plots for Wikibook (en, fr) from data and model. (c),(d),(e) - Words degree distributions plots for Wiktionary (en, fr, de) from data and model. X-axis shows the degree (k) and Y-axis depicts the probability of having a book and word with degree k (p_k).