

10. Life Cycle of Social Networks

Different Wikipedia projects grow differently. This is not surprising, because they are maintained by different communities. Thus they are influenced by different economical, political and cultural conditions. Because each Wikipedia sub project is created in a different language, one can say, that each Wikipedia project represents a different cultural context wherein different topics are important. This view is based on differences in how languages are used and how different cultural aspects are reflected within community driven large projects.

A second perspective exist. On a higher abstraction level, all Wikipedias can be unified, by ignoring the cultural and lingual differences. In this case one can say: Wikipedia is the global encyclopedia. It is a crowd based information storage and knowledge creation system. A growing system with inherent memory. For multiple languages, there exist several Wikipedia instances. All may have comparable properties. If a reasonable approach of normalization of the data exists, one could compare the project life cycles or phases.

To verify this hypothesis we analyze the growth procedure of four Wikipedias. We selected English, Swedish, Dutch, and Hebrew Wikipedias for this sub project. A preliminary report was also published by Schreck *et al.* [?].

10.1. Cultural Aspects of Global Online Networks

The detailed impact of different cultures on Wikipedia content and on the processes, such as content creation and consumption are not in the scope of this work. In our data driven studies we can follow a natural path and differentiate by language. This is easy, because all Wikipedia projects for all languages coexist and they are interrelated already. Each language can define one dimension if a global analysis is desired. An advanced approach uses more specific dimensions. They were derived by Hofstede *et al.* [?] from global survey data. As they show, it is possible to apply factor analysis to determine the predominant cultural dimensions. Hofstede *et al.* defined four cultural dimensions regarding fundamental anthropological problem fields. The dimensions are named: power distance (PDI), individualism (IDV), uncertainty avoidance (UAI) and masculinity (MAS). Long-term orientation (LTO) and Indulgence versus restraint (IVR) were added later. Although this approach is data driven it is not applicable to a global system like Wikipedia. Wikipedia covers many different topics. A clear segregation between cultures is not possible. Many people speak multiple languages. Even if they have a different cultural background, they may contribute to the same Wikipedia project or to a different one, depending on their current background or context. Culture is one context, but obviously not the only one which influences the representation of topics within Wikipedia (see figure 13.4.2 for an illustration of the impact of the lingual context on a topic's representation in different languages). Therefore, Wikipedia seems to be a good source for advanced studies on cultural differences in knowledge formation and knowledge sharing which is related to cultural contexts as well.

10.2. Growth of Social Content Networks

The Wikipedia projects are more than just networks of pages, to which a new page is added at a given time. Pages provide information, they innervate new ideas, lead to questions, and as a consequence, new pages are added by different people. The editorial process can be highly controversial as Yasseri *et al.* [28] and Eckstrand *et al.* [?] show. The system is embedded within user communities which consist of editors and readers. Not all people contribute to Wikipedia, but a critical mass of users seems to be required by a Wikipedia project in order to survive. The evolution of Wikipedia project sizes was already analyzed by Ortega *et al.* [?]. They found that the contributions to Wikipedia are dominantly made by several so called "power users". Based on a calculation of the Gini coefficients for the top ten Wikipedias they state that approximately 90% of all users are responsible for less than 10% of the content. A comparable distribution of user activities in several other wikis - non of them are Wikipedia projects - was found by Stuckman and Purtle [?]. Such a strong bias towards some very active users has to be taken into account. Analysis on the life cycle should not just be build on the editorial activity as presented by Gorgeon and Swanson [?]. They studied the evolution of the topic or concept "Web 2.0" in Wikipedia based on article size, number of editorial actions and number of contributors. As a result, they define four phases for an article: Seeding, Germination, Growth, Maturity (for details see section 5 in [?]). The life cycle phases already take the activity and controversial character of editorial events into account. One can clearly conclude, that editorial activity does not always lead to an increase of content, because higher quality can be achieved by clear statements which are often the result of shorter sentences. Too long articles are sometime misleading or distracting. Different category classes exist in Wikipedia. [?] show, that article size distributions

are bi-modal for English and Polish Wikipedia projects. These studies do not care about the network structure of the articles. Based on the node degree or on a centrality measure one can differentiate between leaf nodes, which contain definitions and well accepted facts and more central pages which are related to many topics and which define context as they aggregate several leaf nodes. Such additional aspects show, that edit activity is not only related to a change of words or sentences. Furthermore, the embedding of a page is important. In many cases it is even not possible to work with just one page, because the selected topic is represented by different pages within the same language. Aggregation over all pages of the topic - or even a full category - and contextual normalization within the local embedding was developed as a part of this work. Such aggregated measures can contribute to advanced life cycle models on the microscopic scale.

A social network is defined by interactions between many people. No matter if the final result is creation of a new resource in a content network or if it leads to a specific temporal state of minds of all connected participants, one can analyze the underlying structure. According to Borge-Holthoefer *et al.* [?] the evolution dynamics of a social community can be described by the size of the giant component, plotted as a function of time. Changes of growth rate can be interpreted as an indicator of existence of such a particular social network which has no physical representation, such as Wikipedia topics. The calculation of the giant component uses already more details, such as the link structure instead of counting words only.

In case of Wikipedia we have a mixture of both. The social ties of interconnect users and editors influence the process of content creation. An edit war is something which goes on without explicitly being announced. But this mental state within the community is measurable and contributes to the life cycle of the articles. Such procedures bind or require energy and information can be lost over time.

To formalize this idea better, I use the concept of Emissivity as an analogy. Although the analogy is weak it helps to understand the many facets within one coherent framework. First, I compare Wikipedia with a physical body. It consists of matter and has a given structure. In Wikipedia we can not find this. Content in digital documents can easily be copied and one has not to care about conservation of mass. Because we try to describe the flow of information, this is not a critical issue as it belongs only to the description. Let's assume we have contribution to a system, like Wikipedia, we can clearly say, the more information it contains, and the better the structure supports easy access to this information the higher the impact or usefulness of it may be. With this in mind I compare Wikipedia with a solid body, which exists in a field of radiation. The incoming energy flow leads to an increase of internal energy and to internal heating. The body emits energy according to its internal state. In an equilibrium state it emits the same amount of energy as it absorbs.

Explain Strahlungsgleichgewicht *Energie wird in z.B. in Form von Strahlung in ein System übertragen. Die Effizienz der Übertragung hängt von vielen Faktoren ab. Vernachlässigt man Wirkungsquerschnitt, Wellenlänge und Pulsform und betrachtet nur die Menge an Energie, die tatsächlich vom System aufgenommen wurde, dann bleibt dennoch zu unterscheiden, welche Form der inneren Energie erhöht wurde. Verschiedene Prozessbeschreibungen oder Modellvorstellungen helfen dabei, solche Situationen zu erklären. Die Erhöhung der Temperatur ist eine recht einfache Vorstellung, die Anregung von Rotationsmoden eine andere, mit einem komplizierteren Modell verbundene.*

Das Ziel dieser Analyse ist es, zu betrachten, ob die Aktivität der Wikipedia Editoren, die sich in Form von Edit Ereignissen zählen lässt, zur messbaren Strukturveränderung des Systems und zum Volumen Wachstum in Beziehung zu setzen ist. Gibt es Phasen, in denen der eine oder andere Anteil dominiert? Wie kann man solche Phasen erkennen?