Modeling the Dynamics of Discussions in Social Networks

Petr Klán, Dept. of System Analysis, University of Economics in Prague, Czech Republic, petr.klan@vse.cz

Abstract: In this paper, a method of modeling the dynamics of electronic discussions is proposed based on the so called FOPDT model (First Order Plus Dead Time) known from the process control. Knowledge of the model points to possibility of estimating dynamic movements in discussions as well as understanding and designing their maintaining and guidance. Real discussions are processed.

Key Words: Discussion dynamics, FOPDT model, process identification

1 Introduction

Social networks are associated with complex systems. In order to investigate properties of such systems, statistical models are widely used. They ignore individual network nodes and prefer average behavior of large number of nodes that carry out communication. Furthermore, average characteristics of communication networks are given by Zipf law [7]. The latter determines, for example, that frequency of Web sites linked from \( k \) other sites (in–links) is proportional to the ratio \( 1/k^2 \). In other words, pages linked from a huge amount of the other Web pages, are very rare. A similar type of dependence can be measured e.g. in the electronic discussion group Chminf–l [5]. Frequency of discussions including \( k \) discussion posts is approximately proportional to \( 1/k \). As shown in Fig. 1, corresponding relationship was measured in the discussion group Chminf–l for year 2010 like in other years. Discussions including more posts than e.g. 15 are rare in this discussion group.

In many cases, the patterns of dynamic behavior in complex systems are given by the size changes in some populations \( n, n > 0 \), described by differential equations of type \( \dot{n}(t) = bn(t)(1 - n(t)) \). The latter results in well known logistic curves \((t \geq 0)\)

\[
n(t) = n(0) \frac{e^{bt}}{1 + n(0)(e^{bt} - 1)},
\]

where \( b \) denotes size change coefficient and \( n(0) \) initial condition [1]. Logistic curves with the characteristic shapes of the letter S represents a common pattern appropriate to describe the dynamics of complex systems, regardless of their origin. In some cases, cumulative dynamics of responses in electronic discussions exhibits the characteristic shape of the logistic curve too, as illustrated in Fig. 2 for Chminf–l discussion carried out on July 22, 2011.

Advanced methods for PID (Proportional + Integral + Derivative) control use mathematical process models which have a potential to study dynamic events related to communication in social networks as well. Primarily, there are three–parameter models [2] or FOPDT models (First Order Plus Dead Time) with the corresponding differential equation \( Ty'(t) + y(t) = Ku(t - L) \) represented by the transient solution \((t \geq L)\)

\[
y(t) = K \left( \frac{-(t - L)}{1 - e^{-\frac{t - L}{T}}} \right)
\]
Figure 1: Illustration of Zipf law. Horizontal axis indicates number of discussion posts and vertical axis frequency of these discussions.

Figure 2: On the S shape of the logistic curve. Horizontal axis represents time, vertical axis shows growth of discussion post population.
given by the unit step response for $y(0) = 0$. An example of such transient response with $K = 27$, $T = 5$ and $L = 1$ is shown in Fig. 3. Variables $y, u$ indicate input and output and non–negative parameters $K, T, L$ gain, time constant and time delay, respectively. Physical units are not specified in this illustrative example. Using Laplace transform, this model corresponds to a transfer function

$$K \frac{e^{-sL}}{T s + 1}.$$

![Figure 3: Three–parameter model step response.](image)

The paper shows that the dynamics of communication networks, under certain conditions, can be identified by the three–parameter models. The latter is yet suitable for purposes of prediction and management of discussions using principles associated with PID control. Since the three–parameter models are widely used in automatic control, various rules have been developed for their use. Here, knowledge of parameters $K, L, T$ is decisive. Just widely known model, the intuitive clarity of its meaning, identification and use allow to better understand the communication dynamics in social networks, including such elements of communication as posting, liking etc. For example, use of the specific value of 63% of $K$ at $t = T + L$ is available for estimating sizes of discussions.

The following section explains motivation for use of three–parameter model for discussion dynamics description. Further sections are concerned with identification of three–parameter models from cumulative data of discussions. Moreover, some real examples of identification in electronic discussion group Chminf–l are introduced.
2 Motivation

The author was a participant in one of the discussions on social network, which debated the topic of possible Wikipedia decline despite the noble goal of Wikipedia to assemble all human knowledge. Although a unique system in the history of civilization, the number of active Wikipedia contributors constantly decreases significantly. The principal reason is considered in constraints that community of Wikipedia contributors has created to discourage new contributors.

Discussion had a classical format. At the beginning it was an initial post about Wikipedia decline giving the link to the paper [8] to which the participants debated. After the end discussion, the author collected the dynamics of discussion, i.e. the cumulative number of posts $y$ to vertical axis versus related times $t$. It results in Fig. 4 where the circle indicates the cumulative number of discussion posts at the time. The first post on the subject appeared after about an hour from the initial contribution (transport delay). It is followed by the relatively steep growth of posts to gradually slowed after 24 hours to reach the final steady state of 27 posts.

![Figure 4: Dynamics of discussion on Wikipedia decline.](image)

If observing dynamics of this discussion then one can not ignore the similarity with the exponential dependence of transient response in Fig. 3. The same transient response with parameters $K = 27$ posts/time, $T = 5$ hours and $L = 1$ hour is inscribed by continuous line in Fig. 4. It well characterizes and describe the dynamics of discussion and associates with a principal question if is it possible to describe discussions by the related transfer function.
or whether the three-parameter models describe the dynamics of electronic discussions in social networks and the Internet in general?

3 Discussion dynamics identification

At the beginning, it is necessary to answer the question whether the dynamics of discussion behaves as a transient response to step change? Yes, it does. The initial post from which the debate unfolds, appears and lasts. Thus, it fulfills the same ‘shock’ role like unit step expressed by $1/s$ in the Laplace transform. At the time of this step or the initial post, it is possible to set up the start time $t = 0$ and (with zero initial condition) to start measurement of transient response as the cumulative number of the posts.

The dozens of discussions were examined to verify if the discussion dynamics captured in Fig. 4 does not represent a random but repeatable shape. For this purpose, the freely accessible archive of electronic discussion group for chemical informatics Chminf–l (https://list.indiana.edu) was selected. The author is a member of this group since 1995 [4]. The group has about 1500 participants and well serve as a model paradigm of communication in social networks because indicating the same patterns introduced by the Zipf law.

Discussion subjected as ‘What are faculty instead of using chemistry reference titles?’ from February 2010 represents a characteristic example. The initial post (unit step) was sent at 16:35 ($t = 0$). The total number of the following posts was 23, which immediately represents gain $K$ associated with the unit step. The dynamics of the discussion is captured in Fig. 5 (irregular curve) together with the unit step response of identified three-parameter model by $K = 23$ posts/time, $T = 5.5$ hours and $L = 0.5$ hour. In order to obtain these parameters, the algorithm of identification described in [2] was used. Therefore, discussion is identified by the transfer function

$$
\frac{23e^{-s}}{5s + 1}.
$$

Discussions with a large number of posts (typically $> 10$) indicate common patterns in the aforementioned public archive. It examines discussions in private social networks, too. Data from large amount of discussions confirm the hypothesis that the dynamics of the discussion can be interpreted (with sufficient accuracy) as the dynamics of the three–parameter model. In other words, dynamics of a social network discussion can be sufficiently represented by three parameters, $K, L$ and $T$.

Moreover, it appears that while the parameters of $L$ and $T$ in the dynamics of the discussions are not too much different (typical delay currently was within one hour and the time constant from 1 to 5 hours), the parameter $K$ representing the cumulative number of posts significantly varies. Thus, the situation relates to dynamic systems with a highly variable gain. On the other hand, the Fig. 4 suggests that discussions with gain of more than $K = 15$ are significantly rare in the investigated discussions.
4 Examples of identification

This section shows examples of identification and illustrates situations met in the real practice of discussions.

**Example 1**: Chminf–l 10.3. 2011, beginning: time 15:59, subject: ISTL article . . . , basic time unit: hour, Fig. 6.

It is a case of the discussion, which had the standard dynamics and that the last two posts were added after many hours. In the case of international discussions due to different time zones (or e.g. working absence), it is a common phenomenon and it currently represents 10% error given by a noise. It is known [3] that, for example, good PID tuning has to deal with 20% changes of the controlled process parameters.

**Example 2**: Chminf–l 6.1. 2012, beginning: time 1:41, subject: Authors listed in search of ACS Pubs, basic time unit: hour, Fig. 7.

The example shows the situation with an extremely large time delay, the process by which the first contribution post to the discussion came after about 12 hours. So long delay is examined in discussions rare and when they do, it is usually given by inadequate time posting of the first discussion post.

**Example 3**: Chminf–l 22.7. 2011, beginning: time 17:23, subject: SciFinder Piracy, basic time unit: day, Fig. 8 Chminf–l 7.11. 2011, beginning: time 20:03, subject: Handbook 92d edition . . . some thoughts, basic time unit: day, Fig. 9.

It is illustration of discussions with long time delays and time constants (in days). Furthermore, Fig. 8 captures a somewhat chaotic shape of the discussion dynamics, in which,
Figure 6: $\frac{e^{-2.5s}}{2.5s + 1}$.

Figure 7: $\frac{e^{-13.1s}}{2.5s + 1}$. 
despite the large variations, it is possible to recognize the exponential character (and finally the logistic curve).

\[ \frac{e^{-1.5s}}{2.5s + 1}. \]

Figure 8: \[ \frac{e^{-1.5s}}{2.5s + 1}. \]

**Example 4:** Chminf–l 22.6. 2011, beginning: time 18:13, subject: Call for papers from Journal of Chemistry and Chemical Engineering, basic time unit: hour, Fig. 10.

Capturing a break dynamics discussion (e.g. due to the absence in work), which may occur in the discussion, unlike the technological process. It is interesting to see that the dynamics of the discussion is interrupted while the delay continues.

**Example 5:** Chminf–l 2010, first 5 discussions including 8 posts, Fig. 11.

Fig. 11 shows examples of the discussion dynamics with a small number of posts. Such discussions are frequently included in the Chminf–l. In one case, the last post came even after more than 30 hours from the initial post. In Fig. 11 a step response without time delay is experienced as well as a delayed response with extremely small time constants as well as response with a time constant of around 10 hours. All of these responses can be identified by three–parameter model similarly as in previous cases. However, using the predictive power of the model, i.e. prediction of the characteristic time \( L + T \) is difficult. On the other hand, the gain \( K \) for newly starting discussions can be estimated by the Zipf law (well introduced eg. in \([7]\)). With each new discussion will be possible, based on the subject of discussion, to estimate its probable order associated with the number of posts.

## 5 Conclusion

The paper shows an innovative use of three–parameter FOPDT models in identifying dynamics of social discussions. By this view, the cumulative dynamics of discussions is described
Figure 9: \( \frac{e^{-0.3s}}{1.5s + 1} \).

Figure 10: Discussion under work absences.
as e.g. dynamics of mechanical systems. The paper gives examples of identification of real discussions in the electronic discussion group called Chminf–l.

It is shown that dynamics of discussions is well described by the three parameter FOPDT model known from the process control. At the same time, the predictive ability of three–parameter models can be used. In the case of a small number of posts, it is possible to use a combination of the model with the effects of Zipf law.

In addition to the prediction and estimation of dynamic properties of discussions, the three–parameter model serves in managed discussions when calculating proportional and integral parts. The proportional part can manifest itself by inserting proportionally opposing views or to removing the offending posts, the integral part then historical experience on the subject, experience from previous discussions, etc.

References

[1] Kahn P.B., *Mathematical Methods for Scientists and Engineers*, J. Wiley & Sons, 1990.

[2] Klán P., Gorez R., *Process Control*, FCC Public, 2011.

[3] Klán P., Gorez R., *Balanced tuning of PI controllers*, European Journal of Control, Vol. 6, No. 6, pp. 541–550, 2000.

[4] Klán P., Mindl J., Štědrý A., Rubešová E., *Chemical Informatics by the Internet*, Academy of Sciences of the Czech Republic, Czech Chemical Society, CHEMagazín, 1999.

[5] LIST.INDIANA.EDU, Mailing Lists https://list.indiana.edu/sympa/home.

[6] Mitchell M., *Complexity: A Guided Tour*, Oxford University Press, 2011.
[7] Pierce J. R., *An Introduction to Information Theory: Symbols, Signals and Noise*, Dover Books on Mathematics, 1980.

[8] Simonite T., *The Decline of Wikipedia*, MIT Technology Review, [http://www.technologyreview.com/featuredstory/520446/the–decline–of–wikipedia/](http://www.technologyreview.com/featuredstory/520446/the–decline–of–wikipedia/), 2013.