Metrics for best practice networks analysis

Pierfrancesco Bellini, Daniele Cenni, Alice Fuzier,
Paolo Nesi, Gianni Pantaleo, Michela Paolucci
DISIT-DSI, Distributed Systems and Internet Technology Lab
Dipartimento di Sistemi e Informatica, Università degli Studi di Firenze, Firenze, Italy
http://www.disit.dsi.unifi.it/, nesi@dsi.unifi.it, paolo.nesi@unifi.it

Abstract

Best practice networks are thematic social networks focused on sharing common area works and goals. The analysis of user behavior on social network is fundamental to tune services and stimulate the network growth. This is even more relevant for best practice collaborative networks where details about collaboration may lead to understand the effective activities and role of people in the several groups. In this paper, a set of metrics for analyzing user behavior related to downloads and other aspects on the best practice network is presented. The metrics and the work presented have been applied, and are currently active, as administrative tools for the Mobile Medicine best practice network.

1. Introduction

Many widespread Social Networks (SN) are mainly focused on contents (e.g., YouTube, Flickr, LastFm) [KruNac08], whereas others are more focused on: establishing relationships among users (Facebook, MySpace, Orkut, Friendster) [NazRaz08]. SN analysis is typically focused on analyzing relationships among users, users and groups, in order to identify which are the most central users and groups, and, on the other hand , which are those that are frozen out, namely being those running the risk of losing interest in the network activities due to a serious lack of involvement. This kind of metrics can be classified in:

- Individual Level Measures: Degree Centrality [Freeman79], Betweenness [Brandes01], Eigenvector Centrality [Newman08], Cohesion Degree [Moody03], etc.
- Dyadic Level Measures: Reciprocity [Mislove08], Assortativity [Yong07], etc.
- Metrics on Network Level: Power-Law Degree Distribution, Connectivity, Fragmentation [Borgatti06], Density [ColeMoré83], etc

On the other hand, Best Practice Networks, BPN, are thematic social networks where smaller groups of users share content, contacts and information for a common goal. Some of them are set up by thematic communities, associations and large companies to capitalize skill and knowledge, by facilitating communications and interactions among personnel and the growth of internal content repository and knowledge. As to BPN, to keep trace and analyze user activities is becoming more and more fundamental for tuning services and predicting needs, such as market trends, reactions of users, the product's acceptance, educational needs and problems, etc.

The main differences of Best practice Networks, BPN, with respect to classical massively widespread SNs may be summarized in what follows, where some of the main peculiar features of the BPNs have been highlighted. Therefore:

- BPNs present a larger percentage of active contributors, for both content and comments with respect to SN, the size of the internal communities are smaller. Users of BPN are typically more motivated to participate since BPN are strongly thematic and connected to the user's work. In SN, the effective/active contributors are typically much less than 10%, all the others are merely observers;
- BPNs are typically without advertising, since the business model is designed to provide a service, in order to understand user needs and it is not meant to make earnings with product advertising on a larger number of users. On such grounds, BPN may have a recommendation system for stimulating cohesion among users and activities and thus taking into account a larger number of facts and activities as depicted in [BelBru08b].
- BPNs have to cope with a large variety of content types and not only with videos and images like it occurs on large SNs. Most SNs can manage only simple audiovisual content (e.g., images or videos, which are the simplest content type to be generated by users, for example via digital camera and/or phones). Many BPNs have to be able to cope with complex interactive content for education, edutainment, and/or entertainment experiences. This means having more complex semantic descriptors for those content items. At

present, there are many multimedia and cross content formats for integrated packaging such as AXMEDIS MPEG-21 [BelBru07], [BelBru08], SCORM/IMS, MXF, NewsML, SportML, ePub, HTML, SMIL, etc.

- BPNs have to cope with a large variety of semantic computing activities/algorithms: for content adaptation and processing (also due to the complexity of content) of user contributions: user generated content, indexing of user contributions/ comments, etc. In BPN, the number of users and items is smaller, so that the complexity of semantic processing in terms of triples and ontology complexity may be larger, though still economically viable;
- BPNs can typically cover more platforms for providing services and content distribution; for example: PC, windows mobile PDAs, iPhone, etc.;
- BPNs have more complex content and services, therefore users/content profiles/descriptors have to consider static as well as dynamic aspects [BelBru08b]. In BPN, like in SN, dynamic aspects are much more relevant since they may be continuously updated. Profiles may be used for advertising and providing recommendations to users on the basis of the skill profile.

This paper focuses on the experience in defining, realizing and analyzing a set of metrics for BPN where a large number of content types, devices and content models is presented (please note that this paper is only marginally on user behavior analysis and semantic computing; a complementary paper can be recovered from [BelBru08b**]).** The experience refers to two BPNs based on cross media content semantic model: the XMF (cross media finder) (http://xmf.axmedis.org) (educational) and the Mobile Medicine (http://mobmed.axmedis.org) (medical and mobile). The proposed solutions have been grounded on AXMEDIS MPEG-21 content model and media content processing language [BelBru06], [BelBru07]. The work presented in this paper, performed on BPN metrics, has been realized to: (i) identify the most appreciated content formats by users, (ii) the user fidelity, (iii) the protraction of interest on the provided content, (iv) the geographic distribution of users, (v) the user attitude in working with other people by accessing cross media complex content via different devices.

The paper is organized as follows. Section 2 reports a short overview on XMF best practice network model with AXCP back office platform for multimedia processing. Section 3 presents a set of metrics to assess user behavior on content access and its related analysis. In Section 4, conclusions are drawn.

2. Best Practice Network Architecture

The BPN model exploiting the cross media multichannel distribution has been created to provide support in distributing cross media content towards different kinds of device: PC, PDA and Mobiles (smarphones, iPhone, iPod, iPad).

The simplified architecture of the BPN considered solution is depicted in Figure 1, where main modules are reported: user management (registration and profiling), mobile support (mobile content distribution and monitoring), user generated content support (in connection with back office tools for content ingestion, processing and adaptation), collaborative support among users (chat, grouping, messaging, notifications, ..), indexing and querying (multilingual indexing, fuzzy support, semantic indexing, ..), recommendation support for content and users [BelBru08b].

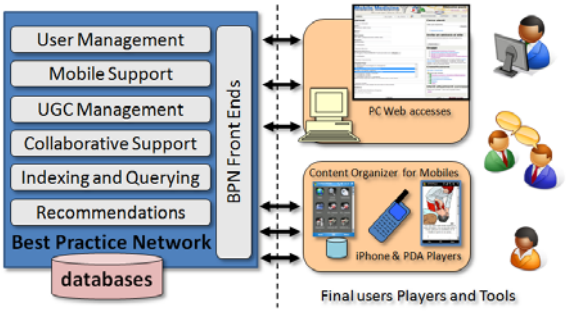


Figure 1 – Simplified BPN Architecture

All the capabilities are deployed and shown to the BPN users via different front-ends, specialized for device category:

- PC (Windows, Mac OS X, Linux, Unix, etc.),
- Smartphone with mobile small footprint web browsers and/or Opera,
- iPhone with a dedicated Application which you can download through and collect content into the local memory for further semantic navigation and querying (see Mobile Medicine iPhone App, http://itunes.apple.com/en/app/mobile-medicine/id359865882).
- Windows Mobile PDA with a dedicated application called AxObjectFinder, which you can download through and collect content on the mobile for further semantic navigation and querying, and getting local recommendations (see http://mobmed.axmedis.org/drupal/?q=en-US/help).

The provided front-ends share the same database and user registration mechanism, and they may provide different services via different sets of languages. For example, the current version of Mobile Medicine is working: (i) with about 20 languages for user interface while only 12 for content; on Windows Mobile a

reduced number of languages is supported on user interface, (ii) the work on user groups is possible on PC and iPhone and it has limited capabilities on Windows Mobile. Moreover, different devices can support different modalities to access certain content: video may be accessed in download, progressive download (also called HTTP streaming), or even via P2P (only on XMF). These capabilities are only available for some devices and solutions.

This is the typical complexity of a multichannel, multi-device, content management and distribution system for a BPN. The designed solution also provides a set of semantic computing capabilities that is only marginally considered within this paper focus [BelBru08b].

2.1 Cross Media Content for BPNs

Most of the massively widespread SN are focused on simple digital items such as single video, audio, images, slides. On the other hand, in BPNs, the set of possible content formats is typically much more complex with respect to the content types adopted in SN. In fact, in BPN content types may range from single files: audio, video, images, documents and animations, simple slides, animations; to cross media such as:

- **interactive guidelines/procedures** to help users to remind the correct procedures and help them in making decisions on the specific cases and pathologies.
- **calculators and tools** which are interactive applications where users may insert the collected data/info so as to obtain from the device/content some estimations/suggestions to be used to take decisions.
- **questionnaires and data collection:** as interactive forms to collect information/data, to perform self assessment, or classical unit-test for training assessment according to Continuous Medical Educational, CME, programs. In some cases, the assessment has to communicate a record of data to the Learning Management System, LMS.
- **sliding show** (sliding images synchronized with audio), typically used for both training activities and establishing also connections with the central server to report actions.

The cross media content and file formats may be of several kinds ranging over: MPEG21, SCORM, ePub, XMF, HTML, SMIL, NewsML, SportML, METS, etc. [BelBru08]. In this context, cross media means objects produced in AXMEDIS MPEG-21 files formats [BelBru07].

3. User Behavior Metrics on BPN

Best Practice Networks organize and classify content, so that users can easily access, retrieve and share. Leveraging Mobile Medicine BPN features; an engine to monitor data has been realized in order to reason upon content descriptors, user profiles, ads profiles, user generated content descriptors, device capabilities descriptors, etc.

Moreover, in order to provide the administrator with an effective tool able to identify the general trends of the BPN including downloads and user behavior, a set of metrics has been defined. Therefore, the identified set of metrics has been implemented, thus allowing the administrator to make his estimation over a selected period of time via simple form and on demand. What follows is a selection of these metrics and the analysis of results about their application on Mobile Medicine BPN. The selected metrics are those that can highlight better the differences between the SNs and the BPNs, on such grounds:

- **S_1:** devices and platforms used to access content: pc-windows (Firefox/ Internet Explorer / Chrome / Opera / Safari) , Linux, Mac Os X, iPhone, etc.;
- **S_2:** content formats which users are interested in, considering: audio, video, document, animations, slides, cross media, etc., with respect to those proposed by the BPN;
- **S_3:** most requested protocols for content access, for example: online play versus content download for further offline play;
- **S_4:** geographical distribution of downloads
- **S_5:** stability of user interests with respect to the distributed content; typically related to typical content metrics such as: the most ranked, the most downloaded, etc.;
- **S_6:** user fidelity, to identify the users who are not only active in a given period, but are more stable in working in the BPN and to check if their activity is related to the number of content they uploaded, they commented, the number of friends they have, etc.

In the next subsection, the detailed description related to the above S_i statistics is reported with corresponding estimation performed on Mobile Medicine BPN in a period of 8 months, starting since August 2009. The data have been depurated from the accesses due to robot and those due to internal testing personnel.

4.1 Download analysis

The metrics used to depict metrics S_1 to S_4 are based on the same reasoning and information, which are reported in the form of histogram or pie.

As to the histogram, the input data are:

$$\begin{cases} x\text{-}axis = t_{Si\_i} = [t_1, t_2, \ldots, t_n] \\ y\text{-}axis = [sum_1, sum_2, \ldots, sum_n] \end{cases}$$

Where: t_i are the possible values of t_{Si_i} variables on which the statistics S_i are calculated:

- $S_{1(devices/platform)}$: $t_{S1_i} = \{$Windows +(Internet Explorer/ Firefox/ Chrome/ Opera/ Safari), Linux, Mac Os X, iPhone, etc.$\}$
- $S_{2(content\_formats)}$: $t_{S2\_i} = \{crossmedia, document, image, audio, video, ...\}$
- $S_{3(access\ mode)}$: $t_{S3\_i} = \{download, online\ play, ...\}$
- $S_{4(geo\text{-}localization)}$: $\begin{Bmatrix} t_{S4\_i,country} \\ t_{S4\_i,city} \\ t_{S4\_i,region} \end{Bmatrix} = \{ISO\ 3166\}$

And sum_i is the number of download per type (t_i). As to the pie chart, it represents the same statistics in percentage.

Usage of Devices and Platform (S_1)

As mentioned, a BPN like Mobile Medicine, offers a range of content formats that may be accessible by different types of devices and platforms. Figure 2 shows the distribution of device/platforms most preferred by users to access content (for both download and direct play).

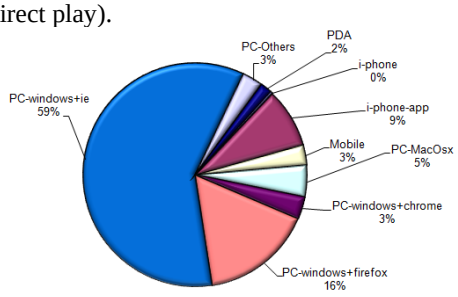


Figure 2: Distribution of content access per device/platform

Most downloads are performed via PC-windows; however, also other operative systems (Linux, Mac Os X) and devices are used (PDA, iPhone, etc.). Therefore, the 12% of users did get access to content by using mobile devices (iPhone, AxObjectFinder), the 86% by using desktop.

Proposed and Preferred Content Formats (S_2)

Given the great diversity of content the platform offers, it was necessary to identify which kinds of format are more appreciated by the BPN users with respect to those provided within BPN. In Figure 2, the distribution of accessed content format is presented. The distribution highlights that cross media content is the most accessed format, followed by video and HTML content pages (still in this content, interactive elements similarly cross media). On mobiles the most selected content formats are cross media with the 82% and video with 9%.

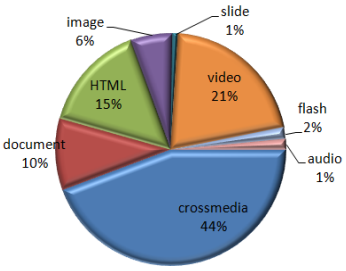


Figure 3: Distribution of accessed content formats

Figure 4 highlights the strong correlation between the content offered population in terms of formats and the actual distribution of content formats as depicted in Figure 2. This graph could estimate what object types are deprecated by users or what classes are more useful for them.

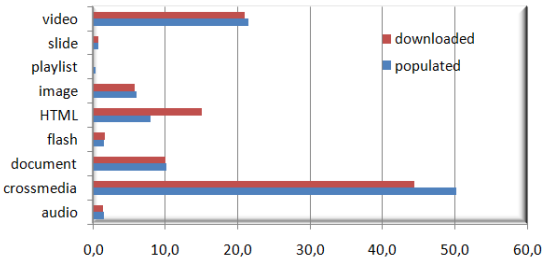


Figure 4: Proposed vs accessed content formats

Content Access Modality of Users (S_3)

The content is accessed via PC and mobiles, while it may be downloaded or directly played via the HTTP connection in progressive download. This last modality is possible on PC and iPhone, while presently it is not viable on Windows Mobile PDA devices endowed with AxObjectFinder tool. As previously discussed only 1% of content is accessed via AxObjectFinder, while 90% is accessed via online progressive download and only the remaining 10% via download (including the accesses via AxObjectFinder). From this analysis, it is self-evident that users prefer to access the content online rather than to have it on their devices. On iPhone, it is possible to have online execution of content or download and successive offline play via the Mobile Medicine App. In this case, iPhone users of Mobile Medicine prefer accessing the video (38%), HTML (26%) and PDF (21%) while only a small part of them prefer other cross media (please note that dosages and medical calculators are classified as

HTML). PDA users prefer accessing the cross media for the 94%. For PDA, cross media include the typical medical tools for dosages, guidelines, checklists, procedures, etc. What is impressive is the adoption of iPhone for on-line video play. It seems to be the most preferred platform for video play.

User Geo-localization (S_4)

The present version of the Mobile Medicine portal is multilingual, with more than 20 languages, while the content is mainly Italian and English. On this basis, we have performed the classical analysis of user geo-distribution to understand the user location in terms of countries, regions and cities. In Figure 5, the country distribution of content accesses is presented; noteworthy is that the greater spread is in Italy, Israel, United States. The distribution of mobile downloads/plays is strongly connected with Italian users (with a 95%) whether compared to the distribution of PC accesses, as reported in Figure 5.

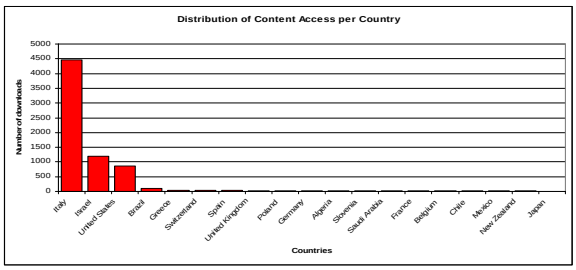


Figure 5: Geo-distribution of downloads

4.2 Content Stability and User Fidelity

Another important factor is the assessment of content stability and user fidelity as described in Section 4.

Object Stability (S_5)

In SN, the distribution of content accesses is typically strongly exponential, with only a small percentage of content highly demanded. In BPN the distribution is still exponential but less explosive and more stable.

On such grounds, an interesting metric would allow to understand which is the level of interest for a given content item within the BPN community, and thus which is the content permanence in the user's interests. To this end, S_5 has been defined to measure the so called Object Stability (see Figure 6a). It measures how many times a content object has been downloaded in a period of time. We have chosen a histogram as the most representative graph type:

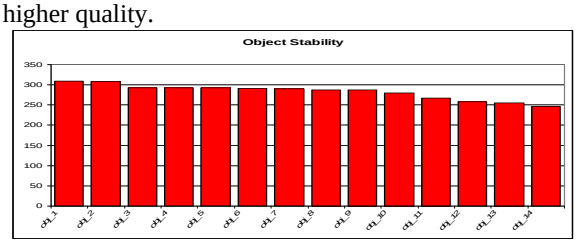
$$\begin{cases} x\text{-}assis = [o_1, o_2, \ldots, o_n] \\ y\text{-}assis = [sum\_o_1, sum\_o_2, \ldots, sum\_o_n] \end{cases}$$

Where: o_i are the object (ids) downloaded and sum_o_i is the number of downloads for each object.

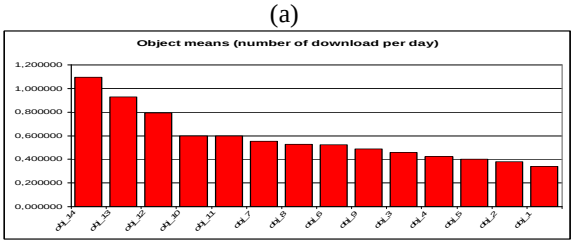
Moreover, it can be also interesting to normalize the distribution, with respect to the duration of the period of interest from first to last download in the period as reported in Figure 6b: date of first (d_f) and last (d_l) download, stability (sum_o_i), mean number of downloads per day ($mean\_o_i$, see Figure 6b) in the period.

Where: $mean\_o_i = \frac{sum\_o_i}{d_l - d_f}$

The comparison of Figure 6a and 6b gives the evidence of the presence of more explosive content, content used more frequently in less number of days (see Obj-14), while others are more stable such as Obj-1. As to the BPN, it is very important to identify both kinds of object, since the most stable one has to provide a higher quality.



(a)



(b)

Figure 6: Object Stability (a) and mean number of download per day in the stability period (b).

User Fidelity (S_6)

The aim of ***User Fidelity*** measure is to assess in a simple shot which are the most active users in the BPN. Even in this case, a more accurate measure would lead to go into the distribution of their action over a time period such as that reported in Figure 7, where the distribution in the last 10 months is presented for a sample of users with different behaviors (some are returning users, other are sporadic, etc.). Simple metrics for assessing user activity can be the simple count of: downloaded and/or played content, connected friends, subscribed groups, preferred content and play lists, uploaded content, different accessed content. In addition to these metrics we have also analyzed the user behavior in terms of number of active days in the period, the so called *user fidelity*. This simple metric is strongly correlated to the number of content: preferred and uploaded (with a

correlation value greater than 75%). The distribution of fidelity has an exponential trend.
In order to assess the user behavior, it can be useful to see which is the user activity in the sections. In Mobile Medicine, the 63% of users access the portal to get/play a single content even if they return back later and in the next days as depicted in Figure 7. Only the 14.3% of them stay for two content items, etc.

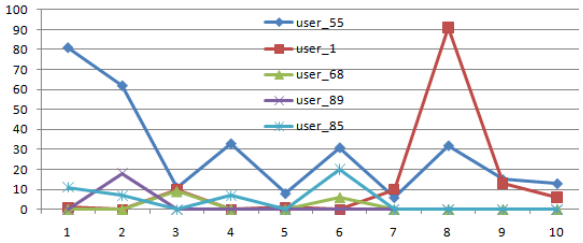


Figure 7: User download distribution in the period.

5 Conclusions

In this paper, a set of direct metrics for the assessment of best practice networks, BPN, has been presented. The analysis of user behavior in BPN is fundamental to tune services and stimulate the network growth. In this paper, a set of metrics related to downloads, kinds of content, different behaviors on PC and mobile, and other aspects implied in BPN has been presented and analyzed with respect to their usage in classical SN. The presented metrics and work have been applied and are currently active, as administrative tools to help the growth of Mobile Medicine BPN http://mobmed.axmedis.org. According to a first validation of the proposal, administrative users did find the direct estimation useful. The proposed metrics have to be tuned according to networks and groups goals. The main discovered issues are related to the different behaviors of users on PC and mobiles, the user preferences with respect to the kind of content, and the identification of the most active users. The proposed metrics can be used as a complementary to the classical SN analysis tools which are more focused on understanding the user relationships, like those mentioned in the introduction.

Acknowledgments:

The authors would like to express their thanks to the many people involved in the validation and to the users of Mobile Medicine BPN; we trust in their understanding for being unable to quote them all in the paper.

References

[BelBru06] P. Bellini, I. Bruno, P. Nesi, “A language and architecture for automating multimedia content production on grid”', Proc. of the IEEE Int. Conf.on Multimedia & Expo (ICME 2006), IEEE Press, Canada, 9-12 July, 2006.

[BelBru07] P. Bellini, I. Bruno, P. Nesi, D. Rogai, “'MPEG-21 for the multi-format and multi-channel content production, protection and distribution”', Proc. of the Int. Broadcasting Conf., IBC 2007, Amsterdam, Sept. 2007.

[Borgatti06] Borgatti, S. – Graph Cohesion, MGT 780: Social Network Analysis, 2006; URL: http://www.analytictech.com/mgt780/2008/slides/graph cohesion.pdf

[Brandes01] Brandes, U. – A Faster Algorithm for Betweenness Centrality, Journal of Mathematical Sociology, Vol. 25, 2001, pp. 163-177.

[ColeMoré83] Coleman, T.F. and Moré, J.J – Estimation of sparse Jacobian matrices and graph coloring Problems, SIAM Journal on Numerical Analysis 20 (1), 1983, pp. 187–209.

[Freeman79] Freeman, L.C. – Centrality in social networks: Conceptual clarification – Social Networks 1, 1979, pp. 215-239.

[KruNac08] G. Kruitbosch, F. Nack. Broadcast Yourself on YouTube – Really? HCC’08, October 31, 2008, Vancouver, Canada.

[Mislove08] Mislove, A., Koppula, H.S., Gummadi, P.K. et al. – Growth of the Flickr Social Network, Proceedings of the first workshop on Online social networks, 2008, pp. 25-30.

[Moody03] Moody, J. and White, D.R. – Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups, American Sociological Review, Vol. 68, No. 1, 2003, pp. 103-127.

[NazRaz08] A. Nazir, S. Raza, C. Chuah. Unveiling Facebook: A Measurement Study of Social Network Based Applications. IMC’08, October 20–22, 2008, Vouliagmeni, Greece.

[Newman08] Newman, M.E.J. – The Mathematics of Networks, The New Palgrave Encyclopaedia of Economics, 2nd edition, Basingstoke, 2008.

[Yong07] Yong, Y.A., Seungyeop, H. et al. – Analysis of Topological Characteristics of Huge Online Social Networking Services, Proceedings of the 16th international conference on World Wide Web, 2007, pp. 835-844.

[BelBru08] P. Bellini, I. Bruno, P. Nesi, M. Spighi, “Tassonomy and review of Complex Content Models”, in Proceedings of the 14th Int. Conf. on Distributed Multimedia Systems (DMS’2008), Boston, USA, 4-6 September 2008, pp.71-76.

[BelBru08b] P. Bellini, I. Bruno, A. Fuzier, P. Nesi, M. Paolucci, "Semantic Processing and Management for MultiChannel Cross Media Social Networking", Proceedings of 2009 NEM Submit "Towards Future Media Internet", 28-30 September 2009, Saint Malo France.