Second International Conference
Modelling and Development of Intelligent Systems
Sibiu - Romania, September 29 - October 02, 2011

Modeling and Analysis for Social Media Network - Case Study: The Small World Type Network for Social Media Networks Analysis in PMML

Ioan Pop

Abstract

The networks of "small world" are exceptional models that can be adapted for modeling networks of social media. This paper proposes a model for social networking model which can be the basis of efficient analysis of these networks. At the end of this paper, we implement the model designed for analyzing social networks. Implementation of the model is done using PMML language (based on XML).

1 Models of social networks

Development potential of Web 2.0 and its synergies with the Semantic Web have facilitated the development of the Social Media. There have been many social networking platforms that favored the development of various social networks operating in various fields of activity of people around the world. A social network is a network of entities that have common interests. Entities in a social network that process information and knowledge are people, groups, organizations, computers, software entities (blogs, software agents, etc.).

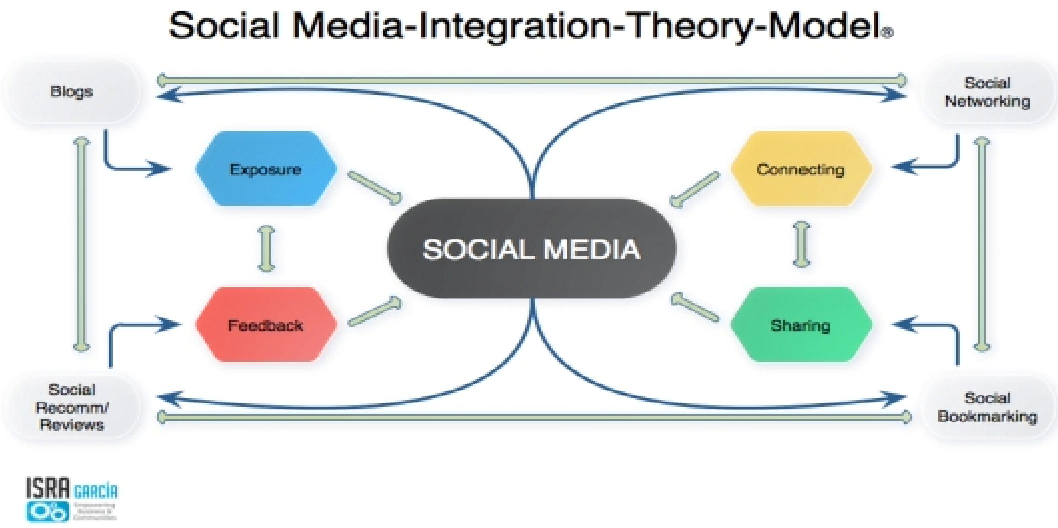

Figure 1.1. Representation of a theoretical model of integration of Social Media [1].

The development of social networking has led researchers to deal with the study and analysis of data stored on social networking platforms to extract knowledge through data mining. An interesting model for social media analysis is illustrated in Figure 1.1.
In social network analysis is used to measure relationships and flows of information between entities that make up the network. Analysis of social networks Social Media enables integration approach. Social Media is seen as an interactive process that makes the exchange of information between network entities with extensive exposure of communication features, feeedback, commitment and sharing. A .

1.1 From random graphs to model social networks

Graphs are used in practice to represent computer science communication networks, organizing information, computer equipment, flow calculation, etc. For example, a website link structure could be represented by a directed graph. Nodes are web pages available on that site and the existence of the directed arc from page A to page B is assured by the existence of a link A to B.

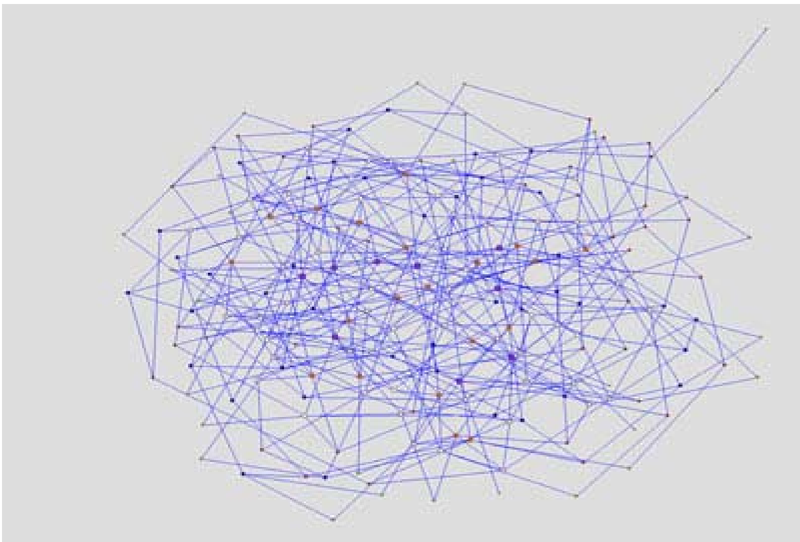

Figure 1.2. A web graph model. [7]

A graph has nodes and arcs. If connections between nodes are determined at random then the graph is called a random graph type. These graphs have special properties such as the probabilities associated with arcs (connections) allocated as contained in the interval [0,1]. An example is given in Figure 1.3.

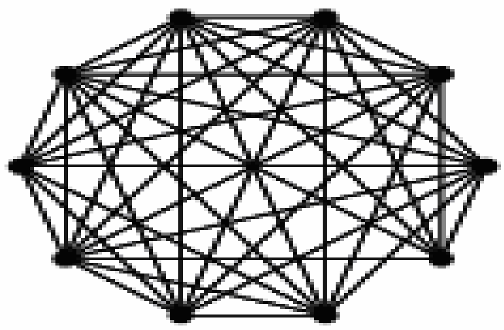

Figure1.3. Type random graph with 10 nodes and uniform distribution of connections [2].

For social network analysis theory of graph model is a powerful modeling tool that helps abstractions needed in this area of exploration. Among the different types of graphs the best type adopted were random graph. [6]. Random graph type is often used to model the web. This model accurately enough plays Web, but in our opinion, is very suitable for modeling social networks.
There are interesting studies regarding fuzzy graphs, a special type of random graphs, graphs which are the subject of the case study in Section 2. Internet Web space can be imagined as a fuzzy graph with the evolution in time. These theoretical models support the activities of the Social Media Study by instrument making analysis offered.

1.2 The power-law in modelling social networks

In general, power-law models scientific phenomena used in fields such as physics, computer science, linguistics, geophysics, sociology, economics. In particular, we refer to the shape of the small world network type.
Regarding the „distribution power law" (briefly called power-law) for random graphs, it has a special application. Power-law applies to a random variable associated with a network of "Small World" described below (section 1.3).
In recent works [3], power-law distributions of type were observed in various aspects of the web. Two ways of modeling work in the social networks are of particular interest to us:
1. First, power laws can be used to characterize the behavior of network users in two interrelated aspects:
a. Statistics of access to social network, which can be easily obtained from server logs (but caching effects) .
b. The power law shows the number of users enjoy access to pages from a particular site. This number is handled and controlled by trees of cache web proxies or clients [4].
2. Second, our context is immediately relevant for the distribution of weights on a tree network graph, ie the probability distribution of the number of connections that go from a full network node (out-degree) or the probability distribution of the number of connections received across the network node (in-degree). The in-degree and out-degree power law-abiding [5].
Power law has the form given by the equality

$$y = \alpha x^{\beta} + o\left(x^{\beta}\right) \quad (1)$$

where α and β are constant, and β is usually called, the scaling exponent, and $o(x^{\beta})$ is a asymptotic function of x^{β} called the "small o"[8]. If social networks, which can be considered special cases of subnets modeled as subgraphs of the web graph, the quantity $o(x^{\beta})$ can be neglected. Logarithms then equality (1) becomes

$$log\, y = \beta\, log\, x + log\, \alpha \quad (2).$$

If we assume that social networks are *y* nodes, each with weight *x*, which satisfy condition (2). It is noted that retains power law form in itself, but re-scaling argument leads to the proportionality constant change. From this we conclude that the law-power frequency model random events that occur in social networks.

1.3 Models based on networks of "small world"

The idea of "small world" was first developed in Milgram's experiment. A "small world" is perceived to be a network of relationships between people which can be identified with each other in this "small world". This refers colloquially to "Six Degrees of Separation", and was the subject of considerable interest in the research community in recent years [9]. According to this concept worldwide, on

average, every person is known by any other person in a large community with information about six steps in relationship knowledge step by step.
Milgram's experiment based on that outlined by the "six degrees of separation" in a small world network type, the conceptualization of this network can become the extrapolation model for social networks of "social media".
It is believed that almost any pair of people in the world can be connected to each other through a short chain of intermediate acquaintances, with a typical length of about six people.
These interpretations have generated new modeling of social networks, which have revealed new features for these social networks.
The mathematical model is based on the notion of random graph associating at each node a weight or a degree of connectivity. The degree of connectivity is given by the probability associated with the event to connect a person with another person.
The most important property inherited from the networks of "small world" is that the graph diameter is relatively small compared to the overall size of the graph associated network. This feature is very important for social networks.
In mathematical terms, the social networks for finding the shortest path between the origin node and the network node of interest in a graph requires, in general $O(n)$ steps, which can be reduced to $O(n \log n)$ for a sparse graph. [10].

2. Implementation of the SWSNM model in the PMML

2.1 SWSNM model based on fuzzy graph with power law.

A fuzzy graph G is a pair defined by $G=(V,F)$, with $V=\{v_i\}$, $F=(f_{ij})$, $0 \leq f_{ij} \leq 1$, where $V=\{v_i\}$ is a set of nodes, $f$ij is the fuzziness of the arc from the node vi to the node vj .[11] This fuzziness of the arc (v_i , v_j) is given by the power law of the probability expressed in the formula (2) from section 1.2.
In figure 2.1 is illustrated a fuzzy graph $G=(V,F)$, where $V=\{v1, v2, v3, v4\}$ with 4 vertices ($y = 4$). Each node in the graph can be connected to another node through a one-way weighted arc with probability given by the power law (2). F is a matrix with 4 rows and 4 columns whose elements are weights corresponding to the graph arcs formed between nodes.

$F =$

	v_1	v_2	v_3	v_4
v_1	1	0.84	0.12	0.84
v_2	0.39	1	0.65	0.71
v_3	0	0.43	1	0.32
v_4	1.00	0.93	0.56	1

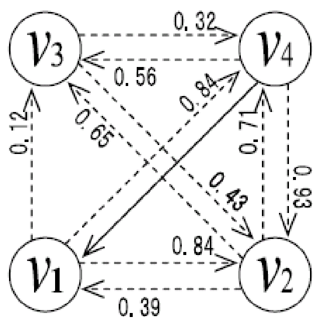

Figure 2.1. Fuzzy graf representation.

The model presented above can be generalized to a social network, which in turn can be seen as a network of the "small world". From this representation, the fuzzy graph G and a social network, can then use a series of measurements of the network arising from the properties associated to the designed graph. The problem of data collection, i.e. discrete values of the weights associated with network connections, is resolved using a simple program that oversees the network of social statistics. Connections made between the attributes of social network members are obtained from server logs and retained in a database.

The attribute's values can be: connectivity of the graph, number of connected components, distribution of nodes per site, distribution of incoming and outgoing connections per site, average and maximal length of the shortest path between any two nodes (diameter), frequency of occurrence of the event connection, connection type, frequency of occurrence of a theme within a connection, frequency of occurrence of a theme within a subgraph of the graph attached to a social network, frequency of occurrence of a theme throughout the term social network, etc..
The model created based on this type of graph characterizing a social network is implemented in PMML

2.1 Implementing in PMML

PMML (Predictive Model Markup Language) is a standard maintained by "Data Mining Group Consortium" based on XML to create data mining models. The structure of any model is described as an XML schema with a specific namespace. A PMML document is an XML tree structure that can represent the model. SWSNM model can be represented in a PMML document where input parameters are weighted connectivity (probability values associated with power law graph model fuzzy). Input parameters (see Fig. 3.1) are represented in a matrix structure as described in Figure 3.2.

```
<xs:element name="Matrix">
  <xs:complexType>
    <xs:choice minOccurs="0">
      <xs:group ref="NUM-ARRAY" maxOccurs="unbounded" />
      <xs:element ref="MatCell" maxOccurs="unbounded" />
    </xs:choice>
    <xs:attribute name="kind" use="optional" default="any" >
      <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="diagonal"/>
          <xs:enumeration value="symmetric"/>
          <xs:enumeration value="any"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:attribute>
    <xs:attribute name="nbRows" type="INT-NUMBER" use="optional"/>
    <xs:attribute name="nbCols" type="INT-NUMBER" use="optional"/>
    <xs:attribute name="diagDefault" type="REAL-NUMBER" use="optional"/>
    <xs:attribute name="offDiagDefault" type="REAL-NUMBER" use="optional"/>
  </xs:complexType>
</xs:element>

<xs:element name="MatCell">
  <xs:complexType >
    <xs:simpleContent>
      <xs:extension base="xs:string">
        <xs:attribute name="row" type="INT-NUMBER" use="required" />
        <xs:attribute name="col" type="INT-NUMBER" use="required" />
      </xs:extension>
    </xs:simpleContent>
  </xs:complexType>
</xs:element>
```

Figure 3.2 Data structure stored in a matrix.
Input parameter values in the matrix are retained in PMML document as shown in Figure 3.3.

```
<Matrix nbRows="4" nbCols="4">
  <Array type="real">1.00 0.84 0.12 0.84</Array>
  <Array type="real">0.39 1.00 0.65 0.71</Array>
  <Array type="real">0.00 0.43 1.00 0.32</Array>
  <Array type="real">1.00 0.93 0.56 1.00</Array>
</Matrix>
```

Figure 3.3. Model input parameters.

When we set input parameters we use the XML schema in a PMML document. The scheme is called "MiningModel". This is coded as shown in Figure 3.4.

```
<xs:element name="MiningModel">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="Extension" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="MiningSchema"/>
      <xs:element ref="Output" minOccurs="0"/>
      <xs:element ref="ModelStats" minOccurs="0"/>
      <xs:element ref="ModelExplanation" minOccurs="0"/>
      <xs:element ref="Targets" minOccurs="0"/>
      <xs:element ref="LocalTransformations" minOccurs="0" />
      <xs:choice minOccurs="0" maxOccurs="unbounded">
        <xs:element ref="Regression"/>
        <xs:element ref="DecisionTree"/>
      </xs:choice>
      <xs:element ref="Segmentation" minOccurs="0"/>
      <xs:element ref="ModelVerification" minOccurs="0"/>
      <xs:element ref="Extension" minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:attribute name="modelName" type="xs:string" use="optional"/>
    <xs:attribute name="functionName" type="MINING-FUNCTION" use="required"/>
    <xs:attribute name="algorithmName" type="xs:string" use="optional"/>
  </xs:complexType>
</xs:element>
```

Figure 3.4. XML Scheme for „MiningModel”.

Different connectivity analysis on a social network can be implemented based on the scheme of Fig. 3.4 "MiningModel" through elements of "segmentation". An element of "segmentation" contains several segments and model combination methods. Each segment contains an element of the predicate that specifies the conditions under which this segment is used. The "segmentation" is coded as shown in Figure 3.5.

```
<xs:element name="Segmentation">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="Extension" minOccurs="0" maxOccurs="unbounded"/>
      <xs:element ref="LocalTransformations" minOccurs="0" />
      <xs:element ref="Segment" maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:attribute name="multipleModelMethod" type="MULTIPLE-MODEL-METHOD" use="required"/>
  </xs:complexType>
</xs:element>

<xs:element name="Segment">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="Extension" minOccurs="0" maxOccurs="unbounded"/>
      <xs:group ref="PREDICATE"/>
      <xs:choice>
        <xs:element ref="ClusteringModel"/>
        <xs:element ref="GeneralRegressionModel"/>
        <xs:element ref="NaiveBayesModel"/>
        <xs:element ref="NeuralNetwork"/>
        <xs:element ref="RegressionModel"/>
        <xs:element ref="RuleSetModel"/>
        <xs:element ref="SupportVectorMachineModel"/>
        <xs:element ref="TreeModel"/>
        <xs:element ref="Extension"/>
```

```
      </xs:choice>
    </xs:sequence>
    <xs:attribute name="id" type="xs:string" use="optional"/>
    <xs:attribute name="weight" type="NUMBER" use="optional"/>
  </xs:complexType>
</xs:element>
```

Figure 3.5. „Segmentation” element.

Segment element is used to combine other models as part of a whole. One way to combine several data processing models to analyze a social network is to specify the attribute "multipleModelMethod" in element segmentation. This method is coded as shown in Figure 3.6.

```
<xs:simpleType name="MULTIPLE-MODEL-METHOD">
  <xs:restriction base="xs:string">
    <xs:enumeration value="majorityVote"/>
    <xs:enumeration value="weightedMajorityVote"/>
    <xs:enumeration value="average"/>
    <xs:enumeration value="weightedAverage"/>
    <xs:enumeration value="median"/>
    <xs:enumeration value="max"/>
    <xs:enumeration value="sum"/>
    <xs:enumeration value="selectFirst"/>
    <xs:enumeration value="selectAll"/>
  </xs:restriction>
</xs:simpleType>
```

Figure 3.6. Elements with attribute analysis "multipleModelMethod".

To use the PMML model for distributed SWSNM results can be posted as a project within the DMG group [12].

3 Conclusions

1. Social networking features deserve to be investigated. These features may show us the functional properties of these networks related to robustness, trend research utility. A relevant modeling of social networks can ensure the functionality of the correlations that exist between content and connectivity. In addition, various thematic communities may become detectable by subconectivity patterns that can be highlighted by a cleverly designed model.
2. The SWSNM can become an effective framework for integrating tools of investigation and analysis of social networks. In this case the model can also lead to better algorithms for the exploitation of th network connections available.
3. The PMML platform provides a number of classes of data mining algorithms that can become useful tools for an efficient analysis of social networks.

References

[1] http://socialmediatoday.com/isra-garcia/278936/social-media-integration-theory-model
[2] R. C. Read, R. J. Wilson, Atlas of Graphs. Oxford, England: Oxford University Press, 1998.
[3] http://arxiv.org/abs/0706.1062.
[4] B. A. Huberman, P. L. T.Pirolli, J. E Pitkow, R. M. Lukose, Strong regularities in World Wide Web surfing. Science 280, 1998, pp.95–97.
[5] S. R. Kumar, P. Raghavan, S.Rajagopalan, D.Sivakumar, A.Tomkins, E. Upfal, Stochastic models for the Web graph, In Proc. 41st IEEE Ann. Symp. on the Foundations of Computer Science, 2000, pp. 57–65.
[6] P. Erdos and A. Renyi, On the strength of connectedness of random graphs, Acta Math. Acad. Sci. Hungar. 12 (1961), pp.261-267.

[7] http://155.198.210.128/~time/networks/ Evolution of Page Popularity under Random Web Graph Models
[8] http://en.wikipedia.org/wiki/Big_O_notationÎn cachePagini similare
[9] S. Milgram, The small world problem. Psychology Today 1, 1967, pp. 61.
[10] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, Introduction to Algorithms, 2nd edn. Cambridge, MA: MIT Press, 2001.
[11] A. N. Gani, K. Radha, On Regular Fuzzy Graphs, Journal of Physical Sciences, Vol. 12, 2008, pp.33-40.
[12] http://www.dmg.org/
[13] J. Scott, Social Network Analysis: A Handbook, SAGE publication inc. California, scd ed. 2000, pp.8-33.
[14] R. Albert, H. Jeong and A.L. Barabasi, Diameter of the World-Wide Web, Nature 401, 1999, pp.130-131.
[15] http://xxx.lanl.gov/abs/cond-mat/9907038.
[16] W. Aiello, F. Chung, L. Lu, A Random Graph Model for Power Law Graphs, Experimental Math Journal, vol. 10, 2000, pp.53-66.
[17] K. Sameena, M. S. Sunitha, Fuzzy graphs in fuzzy neural networks, Proyecciones Journal of Mathematics, Vol. 28, No 3, December 2009, pp.239—252.
[18] http://www.eecs.harvard.edu/~michaelm/NEWWORK/postscripts/webgraph.pdf
[19] A.Z. Broder, S.R. Kumar, F. Maghoul, P. Raghavan, S.Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Graph structure in the web: experiments and models, Proc. 9th WWW Conf., 1999.
[20] R. Kosala, H. Blockeel, Web Mining Research: A Survey, ACM SIGKDD, July 2000. Vol. 2, Issue 1, 1-14.

IOAN POP
"Lucian Blaga" University of Sibiu
Faculty of Sciences
Romania
E-mail: me.ioanpop@yahoo.com