Abstract— Understanding and dealing with the emergent semantics of image and media-based information is one of the most challenging aspects of theoretical, algorithmic, and systems-oriented research in Multimedia. Emergent semantics implies that media is endowed with meaning by placing it in context of other similar media and through factors that are user specific. This means that unlike alphanumeric data, a fixed semantics cannot be assigned to media. While this intriguing property of media-based information has been known for nearly a decade, progress towards development of rigorous frameworks to represent and analyze this phenomenon has been limited. In this paper, we present results that move towards addressing this problem. Specifically, we show how the emergent semantics of a data collection can be first formalized and then captured and represented using network models across users. Using real-world data from a group of users, we then show how such networks can be theoretically characterized and highlight many of their important properties. The primary results communicated in this paper include: (1) a graph-theoretic approach for formalization of the notion of emergent semantics, (2) description of how real-world emergent semantics can be captured and represented as networks, and (3) investigation of the issue of quantitative characterization of emergent semantics through the analysis of these networks.

Keywords- emergent semantics, image and media semantics, human-computer interaction, information modeling, network analysis, multimedia retrieval

I. INTRODUCTION

The notions of syntax and semantics are fundamental to multimedia information systems. Of these two, syntax is perhaps the easier one to define, and is generally understood to be the rules guiding signal or symbol formulation and manipulation. The notion of semantics on the other hand is more complex and generally is interpreted as the meaning associated with a group of symbols. In the context of image and media-based data, in a seminal paper in the area of multimedia information modeling and retrieval [15], Santini et al. posited that meaning is not an intrinsic property of an image but an emergent property of the interaction between a user and an image collection or database within which the specific image resides.

To address the question of associating meaning with an image, a model was proposed in [15], where in response to a query, the database provided information about the entire collection of images in it, rather than just about the few images that satisfied the query. A user would then manipulate the image space directly by moving images around per his/her specific semantic perspective. The goal of the re-organization was to bring together images which the user considered to be semantically related and move away, images that were unrelated. For example, the reorganization could bring closer images that a user considered to be “similar” by a user. To instantiate this idea an exploration interface consisting of three spaces was proposed. These included: (1) the image feature space (2) the query space, obtained by endowing the feature space with a metric derived from the user manipulation of the image space, and (3) the image display and interaction space. Each user-system interaction (defined by some algebra of operators between these three spaces) reconstructed the query space based on the emergent semantics.

Since Santini et al. work, it has been increasingly recognized [4, 16, 17] that the emergent nature of semantics is inherent not only in images, but also in other media and arguably in many sufficiently complex informational entities. For example, it was demonstrated in [4] that the semantics of web-pages can be emergent. This realization has influenced various research directions in multimedia including the development of unified multimedia data models [16], experiential computing [5, 11, 17, 18] and analysis of user-data interactions on the web in manners that directly account for user specificity [4, 19]. Yet, our understanding and ability to deal with emergent semantics remains nascent. For instance, at the current state-of-the-art, neither formal models nor methods for rigorous characterizations of emergent semantics exist. Further, we are as of yet unable to answer questions about how the emergent semantics of a particular user evolves over time or how the emergent semantics of the same data set from the perspectives of different users may vary or be related. Given this context, the key contributions of our research include:

- **Formalization of the notion of emergent semantics**: We propose a graph-theoretic framework to formally describe and represent the emergent semantics associated with a collection of media (in our case, images). Specifically, this framework allows us to capture and represent the different “meanings” one or more users ascribe to subsets of (a given) data collection over time.

- **Application to real-world emergent semantics data**: We show how the proposed theoretical framework can be
used to capture and represent the real-world emergent semantics of an image collection both longitudinally across time and (simultaneously) across different users.

- **Quantitative characterization of image semantics**: using the data from the above step, we demonstrate how the emergent semantics of a given data collection can be quantitatively characterized and analyzed. Specifically, we investigate the potential of using established theoretical network models to describe the emergent semantics of a data collection using a number of local and global network properties.

While our current research explores the fundamental aspects of modeling emergent semantics, the possible practical applications of a rigorous understanding of emergent semantics are many. As we shall see in Sections IV-V, the approach proposed by us allows comparing the emergent semantics associated with a data collection by different users in terms of the underlying mathematical models. Such information can be used in personalization or in identifying users who have similar perspectives on the data. This can lead to new retrieval paradigms and be of use in recommendation systems. Deeper analysis of the emergent semantic network of a data set (see Section II for the theoretical foundations) may also be valuable in designing graph-databases for such information, for distributed storage, and for designing retrieval protocols in such systems.

II. FORMAL REPRESENTATION OF EMERGENT SEMANTICS

The nature of emergent semantics implies that the meaning of image or media-based data can only emerge from user-data interactions. The full meaning attributed to an image thus depends not only on its signal-level components but also on the user specificities and state at the time of the query as well as on the related images in the database. Systems such as those described in [11, 15, 17], have tried to include the influence of emergent semantics by allowing users to organize the data so as to reflect their own interpretation of relevant similarities. We build upon this idea to capture the emergent semantics associated with an image collection as follows: instead of expressing relationships between images by moving them around as done, for instance, by Santini et al., we empower users to hyperlink images (and specific regions between images, if they so desire), to reflect the particular semantic relationship that relates these images. Furthermore, each hyperlink is annotated to describe the particular meaning (or relationship) a user ascribes to the images. Given an image collection, over time, a user would presumably define different relationships over different subsets of the images. A user could also ascribe different meanings to the same subset of images by varying the annotations of the corresponding hyperlinks. In the following we formalize this idea and show how it can be used for representing the emergent semantics associated with a specific data set (possibly across multiple users over time).

Let \( \mathcal{M} = \{m_1, m_2, \ldots, m_n\} \) denote the set of all images in a database. Let also the power set of \( \mathcal{M} \), consisting of all subsets of \( \mathcal{M} \) be denoted by \( \mathcal{P}(\mathcal{M}) \). In the following we will denote by \( \mathcal{P}(\mathcal{M})_i \), the \( i \)th element (subset) of \( \mathcal{P}(\mathcal{M}) \). Finally, let \( \Psi \) denote a set of labels or terms. We define the interpretation structure \( I \) of \( \mathcal{P}(\mathcal{M})_i \), as the pair:

\[
I(\mathcal{P}(\mathcal{M})_i, \Psi) = (\mathcal{P}(\mathcal{M})_i \times \mathcal{P}(\mathcal{M})_i, \Psi) \quad (i=1,2, \ldots, n) \quad \Psi \in \mathcal{P}(\mathcal{M}), \quad \Psi \in \Psi
\]

The interpretation structure can be thought of as a graph, nodes of which are images drawn from some subset of \( \mathcal{M} \), and whose edges represent some relationships between the respective images. In our formalism, these relationships are described using labels \( \Psi \) from \( \Psi \). Now, consider the \( r \)th query (information goal), as a consequence of which the user interacts with the data and re-organizes it to reflect his/her semantic interpretation at that point in time. Let \( \Phi_r \) denote the (re)organization induced by the user in response to this query. Since \( \Phi_r \) is but the imposition of certain relationship(s) on some subset of \( \mathcal{M} \), it can be described, in terms of an interpretation structure as shown in Eq. (3), where the subscript \( r \) serves the role of an index and denotes that the interpretation structure corresponds to the \( r \)th query (information goal):

\[
\Phi_r = I_r(\mathcal{P}(\mathcal{M}), \Psi)
\]

The central idea behind emergent semantics is that for a different query on the same data, a user may impose another interpretation structure on the very same subset of images or involve new images from the database, possibly with yet another interpretation structure. Therefore, the totality of emergent semantics, denoted by \( \Phi \), for a particular data set \( \mathcal{M} = \{m_1, m_2, \ldots, m_n\} \), can be described by the union:

\[
\Phi = \Phi_1 \cup \Phi_2 \cup \cdots \cup \Phi_n
\]

In Eq. (4), \( \Phi \) represents the interpretation structure induced by the \( i \)th query. Thus, the emergent semantics associated with the data is the union of all the interpretation structures applied to the subsets of the data. We represent the result of the union operation on interpretation structures using a labeled multigraph \( \Gamma_r \), which is defined as an ordered pair \( \Gamma_r(V, E) \), where \( V \) is a set of vertices (corresponding in our case to the images/media elements), \( E \) is a multiset of unordered pairs of vertices and associated with each pair of vertices \( (u, v) \) is an labeling function \( \lambda : (u, v) \rightarrow \Psi \). We remind the reader that a multigraph can have multiple edges between the same terminal vertices. In \( \Gamma \) therefore, the labels assigned to each edge correspond to elements of a specific interpretation structure. Given an image collection \( \mathcal{M} \) and its total emergent semantics \( \Phi \), in the following we shall denote the multigraph representation of \( \Phi \), by \( \Gamma(\Phi) \), and call \( \Gamma(\Phi) \) the emergent semantic network (abbreviated henceforth as ESN).

A. Example

Consider the image collection \( \mathcal{M} = \{m_1, m_2, \ldots, m_{10}\} \) shown in Figure 1 (top) consisting of popular tourist destinations in California. Let the set of labels be \( \Psi = \{"peaks", "nature", "man-made", "paintings", "bay-area", "central-\}

...
The interpretation structures corresponding to these labels are:

\[ \Phi_1 = I_1(\{m_1, m_5, m_7, m_9, m_{10}\}, \psi = "peaks") \] (5)

\[ \Phi_2 = I_2(\{m_1, m_3, m_5, m_6, m_9, m_{10}\}, \psi = "nature") \] (6)

\[ \Phi_3 = I_3(\{m_2, m_4\}, \psi = "man-made") \] (7)

\[ \Phi_4 = I_4(\{m_2, m_3\}, \psi = "paintings") \] (8)

\[ \Phi_5 = I_5(\{m_1, m_2, m_7, m_9\}, \psi = "bay-area") \] (9)

\[ \Phi_6 = I_6(\{m_3, m_5, m_8, m_{10}\}, \psi = "central-California") \] (10)

The multigraph corresponding to the totality of the emergent semantics for this example is shown in Figure 1 (bottom). The reader may note that the labeled multigraph indeed captures the emergent semantics of this image collection. For instance, consider the variability of semantics associated with the image \(m_1\).

III. CAPTURING AND REPRESENTING ESNs

We build on the notion of multi-faceted hyperimages proposed by us in [18] to develop a system, based on user-data interactions, for determining and storing the ESN of a set of images. A hyperimage is an image that supports hyperlinks between its regions and regions of other images in a manner akin to web-pages. Facets refer to categories that are used to characterize information in a collection and are realized as a set of labels. In a multi-faceted hyperimage collection, each hyperlink can have multiple facets. Multi-faceted hyperimages can be used to represent the emergent semantics of an image collection as defined by Eq. (4). This is because hyperimages allow us to represent the multiple semantic relationships between images through the hyperlinks. Furthermore, these semantic relationships can be described by using the facets associated with each links. Finally, the ability to link individual regions of images allows the additional advantage of constructing highly precise definitions of interpretation structures. We have built a system that implements this idea. This system allows the user to choose two images and link possibly specific regions within them based on some semantic notion. Next, a facet (label) describing the relation between the selected regions/images is assigned. We note that the facet can be an arbitrary description as decided by the user. The link and facet information is stored and can be utilized for browsing, processing, or analysis. This system also supports browsing the data; starting from an image that is hyperlinked, one can browse the image collection by following the hyperlinks in a manner similar to web-browsing. In Figure 2, we present a network representation of the ESN of an image collection captured and represented using this system.

Figure 1. (Top) Collection of ten images displaying landmark locations in different parts of California. The landmarks \(m_1\), \(m_2\), \(m_5\), \(m_7\), and \(m_9\) are in the bay area, landmarks \(m_3\), \(m_4\), \(m_6\), and \(m_{10}\) are in central California. (Bottom) The multigraph representation corresponding to the interpretation structures shown in Eq. (5) - Eq. (10). Each vertex in this graph corresponds to an image and each edge is created as part of an interpretation structure and represents some relationship between the corresponding vertices (images).

Figure 2. Hyperimage network depicting the emergent semantics network (ESN) of an image collection. The network was built over time as the user hyperlinked/browsed the image collection from a variety of different semantic perspectives.

IV. QUANTITATIVE ANALYSIS OF EMERGENT SEMANTICS

We focus on two key goals: (1) Determining network properties that can be used to characterize specific topological aspects of ESNs. (2) Determining if rigorous mathematical models can be used to describe the ESNs. As described in the previous section, the emergent semantics for a data collection can be described using a network representation where each edge corresponds to a semantic interpretation of the two regions/images it connects. This opens the possibility of analyzing emergent semantics and answering the above questions using techniques for network comparison and
The analysis in this paper focuses on analyzing the topological characteristics of ESNs. That is, we consider only the connectivity characteristics of the ESNs (and not the labels). The reason that such a purely topological analysis has value is because there is a one-to-one mapping between the semantic interpretations (labels) and the edges in an ESN. In general, such an analysis is complicated due to the NP-complete nature of the underlying subgraph isomorphism problem. Consequently, network analysis methods typically rely on comparing local or global network properties to facilitate analysis. In the following, we define the properties used in our investigations and elaborate on the problem of finding models to characterize an ESN.

A. Overview of Local and Global Network Properties

A number of global and local properties have been proposed in literature to analyze networks. Due to space constraints, in the following, we briefly review both global and local network properties and refer the reader to [6, 7, 9, 20] for detailed expositions.

Given a network \( G(V,E) \), common global properties include:

1. **Average Degree**, defined as the average of all out degree of vertices in the network. A related characteristic is the **degree distribution** describing the probability of a node having degree \( k \).
2. **Average Network Diameter** computed as the average of all-pair shortest paths in the network. (3) **The shortest path-length spectrum** describing the distribution of all-pair shortest paths.
3. **The clustering coefficient**, indicating the tendency of the network to have highly interconnected subsets of nodes.
4. **The clustering spectrum**, defined as the distribution of average clustering coefficients of all nodes of degree \( k \). To illustrate how the global properties can be used, note that the degree distribution for Scale-free networks is known to follow the power-law while in the case of Erdős-Rényi random graphs, the degree distribution is Poisson. Similarly, the average network diameter of Erdős-Rényi graphs is proportional to \( \log n \), while for real world scale free networks (degree exponent \( \gamma : 2 < \gamma < 3 \)), the diameter is proportional to \( \log \log n \).

Unlike the global network properties, local properties capture the local structural similarity between two networks. One strategy lies in identifying subgraphs that appear in the network much more frequently than in a random network. We propose to utilize a related but different approach towards characterization of networks using the concept of graphlets proposed in [6]. This analysis is based on identifying small (3-5 node) connected non-isomorphic induced subgraphs, called graphlets, of the ESN \( \Gamma(\Phi) \). The reason we (like other investigators) limit the analysis to 3-5 node graphlets is because the number of graphlets on \( n \) nodes increases super-exponentially with \( n \) and also because graphlets even on small number of nodes have been proven to provide powerful insights about the local structural characteristics of complex networks [6, 13]. We note that there are two possible graphlets on 3-nodes, six graphlets on 4-nodes, and twenty-one graphlets on 5-nodes, leading to a total of twenty-nine graphlets on 3-5 nodes. For an illustration of the topology of these graphlets, the reader is referred to [6].

B. Mathematical Modeling of ESNs

Our approach for finding mathematical models that can describe the networks \( \Gamma(\Phi) \) is based on two directions of investigation. As part of the first direction, we analyze the global characteristics of ESNs in terms of the five properties described above. The second direction involves model fitting in a generative sense. That is, we postulate that a correct (well-fitting) model would generate graphs that resemble the structure of the emergent semantic graph being analyzed across a range of global and local network statistics. In the following, we discuss each of these directions.

In computing the global properties, we note that in the hyperimage-based representation of an ESN, a user may connect images or regions between images. Furthermore, the user may identify semantically different regions within an image. The semantic variability associated with an image can thus be reflected through: (1) multiple edges connecting an (entire) image to other images, or (2) edges connecting different regions of an image to other images or regions of other images, or (3) regions within an image. Now, consider a measure such as the average degree of the ESN. This measure is based on computing the out-degree of the nodes in the ESN. Accounting for the multi-edges of a vertex is important here since they indicate the diversity of semantics associated with the underlying image. Further, regions within an image can be thought of as self-edges and should also be accounted for. In contrast, computation of characteristics like the average network diameter, shortest path-length spectrum, clustering coefficient, and clustering spectrum are based only on the consideration of connectivity between vertices and exclude consideration of self-edges.

The average degree \( \Delta \) of an ESN \( \Gamma(\Phi) \), in light of the above discussion, is computed as follows:

\[
\Delta = \frac{1}{K} \sum_{i=1}^{K} \sum_{j} (r_j + d(r_j))/K 
\]  

(11)

Where, \( r_i \) denotes the regions in an image \( I \in \Gamma(\Phi) \), \( d(r_j) \) denotes the number of edges (hyperlinks) connecting a given region of \( I \) to other regions or images in \( \Gamma(\Phi) \) and \( K \) is the number of images in \( \Gamma(\Phi) \) to which a user has associated a semantic interpretation either through hyper-linking or through identification of regions (self-linking). To compute the other characteristics that are purely based on connectivity, we transform the multigraph \( \Gamma(\Phi) \) into a simple derivative graph \( \Gamma'(\Phi) \), where multiple edges between vertices are replaced by a single edge and self-edges are ignored prior to computing the global properties (2) – (5) described above in Section A.

Prior to discussing the second direction of our investigation, which involves model fitting, we point to the reader that the question of determining such models is a fundamental one. This is because of the following reasons: first, finding the best fitting model is self-evidently one of the most important ways...
to rigorously characterize emergent semantics networks. Second, appropriate models, once determined, can be central for comparison of emergent semantics across users and in characterizing its time-dependent evolution. For instance, we can determine if the emergent semantic graph of different users on the same data set belong to the same graph family or to different families. Similarly, such models can allow us to rigorously analyze the evolution of the emergent semantic graph for a specific user across time. Third, from the perspective of network characterization and system development, determination of an appropriate null model is crucial, since an inappropriate model can lead to errors, for instance, by identifying overrepresented/underrepresented subgraphs that would be ignored under a correct model.

In this paper, we consider an empirical fitting strategy to investigate this problem. That is, we take different data sets that are obtained by capturing user-behavior in real-world settings and consider a series of established network models to answer the question of finding appropriate model(s) for the emergent semantics in the data. The network models considered by us span the spectrum from classical models, like the Erdős-Rényi random graph [3] (abbreviated henceforth as ER-model) to recently proposed ones like the stickiness model (STICKY) [12]. In addition to these two, the list of models considered by us include: random graphs matching the degree distribution of the real-world network being analyzed (ER-DD) [8], Barabasi-Albert model producing networks with scale-free degree distribution (SF) [1], and n-dimensional geometric random graphs (GEO-nD) [10]. We determine the fit of these models to our data by computing the global and local properties of the ESNs obtained from different users and comparing them with each of the aforementioned network models. Such an empirical approach is commonly used in analyzing real-world networks. However, certain provisions need to be made while interpreting the results from such an approach. First, it should be noted that any characterization of the emergent semantics will necessarily be a function of the experiment design, data set(s), and specific users. Second, while the aforementioned list of models includes many of the most important network models, it is possible, that a model that has not been considered in the above list, may yield the best fit. Consequently, the notion of the “best” model is an empirical one, subject to the list of models analyzed.

Unlike comparisons based on global properties, the comparison of local properties requires the definition of an appropriate similarity measure. Recall, that graphlets quantify the local connectivity patterns of the network. Thus, a natural way to characterize a network is to consider the frequency of each of the 29 possible graphlets that can occur on 3-5 nodes. Two networks can then be compared by defining a measure on the graphlet frequency of the networks being compared. We use a measure called the relative graphlet frequency distance (RGF-distance) [13], which is based on this idea. Given two ESNs Δ and Ω defined on a set of images, the RGF distance is denoted by $d_{\text{RGF}}$ and is defined as:

$$d_{\text{RGF}}(\Delta, \Omega) = \sum_{i=1}^{29} | f_i(\Delta) - f_i(\Omega) |$$  \hspace{1cm} (12),

where

$$f_i(\Gamma) = -\log(N_i(\Gamma)) / \sum_{i=1}^{29} N_i(\Gamma)$$  \hspace{1cm} (13)

In Eq. (13), $N_i(\Gamma)$ denotes the number of graphlets of type $i \in \{1, 2, \ldots, 29\}$ and the logarithm is utilized to ensure that the distance measure is not dominated by the most frequent graphlets. Since it is not possible to directly apply this analysis to multigraphs, we factor the ESN $\Gamma(\Phi)$ into another derivative graph which we shall denote by $\Gamma^{n}(\Phi)$. In this factoring process, each vertex of $\Gamma(\Phi)$, which corresponds to an image, is replaced by a set of vertices in $\Gamma^{n}(\Phi)$, with each vertex in this set corresponding to an identified region of the original image. The edges in $\Gamma^{n}(\Phi)$ represent connectivity between regions. Since none of the users we studied, assigned multiple edges to the same terminal regions, this process essentially converted the multigraph $\Gamma(\Phi)$ to a simple graph without excluding any information.

V. EXPERIMENTS AND ANALYSIS

Our experiments were designed to investigate the following questions/issues: (1) Analysis of global network characteristics: We analyze the ESNs ($\Gamma(\Phi)$ or $\Gamma^{n}(\Phi)$) to determine the global properties described earlier. The ESNs used by us were obtained through a multi-user longitudinal study (see following subsection for details). (2) Null model determination: This question relates to the problem of determining the theoretical model which best describes a real-world ESN. We addressed this question by generating networks based on the five network models described above (ER-model, STICKY-model, ER-DD model, SF-model, and GEO-nD model) and computing the RGF-distance between the factored ESN $\Gamma^{n}(\Phi)$ and the generated models. (3) Distinction of emergent semantics across users: Using the data obtained from our user study, we compare the models for different users and analyze how the emergent semantics induced by different users on the same data set are related in this sense. (4) Evolution of emergent semantics for individual users: For this question we investigate if there are any changes in the theoretical models that best describe the factored ESN $\Gamma^{n}(\Phi)$ of specific users over the duration of the study.

A. Data Generation

We generated a data sets (henceforth called data-set I) consisting of 500 images of various points of public interest in San Francisco. Next, a longitudinal user study was carried out involving 10 users. Each user was asked to independently interact with the data using the system described in Section 3 with the goal of organizing the images from any number of semantic perspectives they desired. The ESNs that were created by each user were captured by us once every two weeks over the duration of eight weeks. At the end of the study period, the networks arising from the data sets were
evaluated by us in consultation with the corresponding user to determine how much the user interacted with the data and how well they related to the images. For each ESN, the bulk of the network analysis was conducted using the GraphCrunch [6] software package.

Figure 3. Fit of the five network models to the ESNs of users 1, 4, and 6 (left to right) in terms of global properties. Top row: the average network diameter. Middle row: the clustering coefficient. Lower row: spectrum of shortest path lengths.

**B. Global Network Characteristics**

In Table I, we summarize the cumulative number of images over the four time periods with which each user interacted (in terms of assignment of relationships). The table also presents the number of semantic interpretations (regions and/or links) assigned by each user to the subset of images considered by them. The fourth column of the table shows the average number of regions per image for each user. The final column of the table shows the average degree distribution for each user computed using Eq. (11). As can be seen, for every user, the average degree of the corresponding \( I(\Phi) \) is greater than 2, with the minimum value of \( \Delta \) equaling 2.21 and the maximum value of \( \Delta \) being 6.56. That is, on an average, every user had more than two semantic interpretation associated with each image they considered. It should be noted here that the number of images considered by each user from the total set of 500 images is relatively small. In follow-up discussions conducted with the users, it emerged that some users had only interacted with a small subset of images, as they did not have any experiences they could associate with the other images. The use of personalized image collections may ameliorate this problem. From an analysis perspective, a small number of images may imply that certain statistical network properties may not fully emerge. Studies where users are directed (either through purposive directions or through requirements of generating a certain number of semantic interpretations of the data) can be used to create and study larger ESNs. In this context it is important to point out that our choice of experiment design was driven by the goal of obtaining an unbiased picture of how users assign meaning to images, which is difficult in directed studies.

In Fig. 3, we present the fit of the network models considered by us in terms of three global properties: average diameter, clustering coefficient, and the spectrum of shortest path-lengths. For the average diameter and the clustering coefficient, the absolute difference of the data and the various models are shown on the Y-axis. For the shortest path-length spectrum, the Pearson correlation between the data and the models is shown on the Y-axis. The four time points at which the measurements were made are shown on the X-axis.

Figure 4. Average fit of the five network models to the real world emergent semantic networks arising from data-set 1 consisting of 500 images. The ER-DD model [8] and the STICKY model [12], provided the two best fits (smallest \( d_{ERR} \)). In the above graph, the time intervals are on the X-axis and the \( d_{ERR} \) values on the Y-axis.

<table>
<thead>
<tr>
<th>User</th>
<th>Number of Images</th>
<th>Number of Regions</th>
<th>Avg. Number of Regions</th>
<th>Number of relations</th>
<th>Average degree of connectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>135</td>
<td>145</td>
<td>1.07</td>
<td>174</td>
<td>2.36</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>41</td>
<td>1.24</td>
<td>69</td>
<td>3.33</td>
</tr>
<tr>
<td>3</td>
<td>93</td>
<td>159</td>
<td>1.70</td>
<td>161</td>
<td>3.44</td>
</tr>
<tr>
<td>4</td>
<td>201</td>
<td>244</td>
<td>1.21</td>
<td>561</td>
<td>4.00</td>
</tr>
<tr>
<td>5</td>
<td>66</td>
<td>70</td>
<td>1.06</td>
<td>78</td>
<td>2.24</td>
</tr>
<tr>
<td>6</td>
<td>47</td>
<td>59</td>
<td>1.25</td>
<td>60</td>
<td>2.53</td>
</tr>
<tr>
<td>7</td>
<td>23</td>
<td>24</td>
<td>1.64</td>
<td>23</td>
<td>2.21</td>
</tr>
<tr>
<td>8</td>
<td>75</td>
<td>80</td>
<td>1.06</td>
<td>412</td>
<td>6.56</td>
</tr>
<tr>
<td>9</td>
<td>33</td>
<td>42</td>
<td>1.27</td>
<td>47</td>
<td>2.69</td>
</tr>
<tr>
<td>10</td>
<td>38</td>
<td>46</td>
<td>1.21</td>
<td>84</td>
<td>3.42</td>
</tr>
</tbody>
</table>

**C. Null Model Identification**

To find the best fitting null model for a given factored ESN \( I''(\Phi) \), we generated 50 random networks for each of the ER, ER-DD, GEO-nD, SF, and STICKY network models. Subsequently, each of the ESNs (at weeks 2, 4, 6, and 8) were
compared with these model networks using the RGF distance. In Fig. 4, we present the fit of the models to the averaged data. As can be seen, the ER-DD model [8] was found to provide the best fit on average with the STICKY model [12] providing the second best fit to the data. A detailed look at this graph shows that the ER-DD model provided the best fit across all time points. However, the fit provided by the STICKY, SF-BA, and ER models was close for the first (2-weeks) and second (4-weeks) time points. At 6-weeks, the STICKY model provided a clear second-best fit and at 8-weeks, the STICKY and SF-BA models were indistinguishable. Across all time points the 3D geometric random graph (GEO-3D) [10], provided the weakest fit.

Figure 5. Model variation across users and time for users 1-6 (in row-major order) for data-set I. See Figure 5 for the data on users 7-10. The ESNs in this case were constructed over a data set consisting of 500 images. The number of vertices (regions) for each user is presented in Table I. The RGF-distance is used for model fitting. The abscissa shows the four time points at which measurements were made.

D. Model Variation Across User and Time

In Fig. 5-6, we present the data obtained from fitting various models to the factored ESNs of ten users for data-Set I. Two important observations that can be made are: first, we can observe variations in the best fitting models from one user to another. For instance, the ER-DD model was found to provide the best fit for users 1, 3, 5, 6, 7, and 9. In the remaining cases, the SF-BA model was found to be the best for user 2, the STICKY model for users 4 and 10, and for user 8 the STICKY, ER, and GEO-3D models provided the best fits for different points in time. Second, for almost all users, a single best-fit model could be obtained in the longitudinal setting. The only exception to this observation was user 8 for whom the STICKY and ER models were best fitting at weeks 2 and 4 respectively. At weeks 6 and 8, however, the Geo-3D model provided the best fit. This observation seems to indicate that while there are best-fit model variations between users, model variation across time for the same user is small. Furthermore, the best fit can almost always be clearly distinguished from the second-best fit in terms of the RGF distance.

I. CONCLUSIONS

In this paper we have investigated one of the most fundamental characteristics of image-based data, namely, their emergent semantics. We have shown: (1) how the notion of emergent semantics can be formalized, (2) how it can be captured and represented as networks, and (3) conducted early investigations on the question of rigorous and quantitative characterization of emergent semantics. The implication of these findings for the field is significant. For one, it will allow researchers to pursue the development of multimedia retrieval systems using not just anecdotal notions, but based on solid theoretical underpinnings. We believe that a crucial contribution of this paper is the development of a well defined framework within which such questions can be studied further.

Our research has also opened up many new questions in both multimedia and network analysis. Included amongst these are the possible development of theoretical network models specifically for emergent semantic and the study of features and metrics for characterizing and comparing ESNs outside those investigated in this paper. Also of specific interest in this context is the development of network analysis methods that can directly be applied to labeled multigraphs such as those induced by the ESNs.

REFERENCES


