

# KANDINSKY: Abstract Art-Inspired Visualization of Social Discussions

Christina Lui  
Nanyang Technological University  
Singapore  
lu0003na@ntu.edu.sg

Sourav S Bhowmick  
Nanyang Technological University  
Singapore  
assourav@ntu.edu.sg

Adam Jatowt  
Kyoto University  
Japan  
adam@dl.kuis.kyoto-u.ac.jp

## ABSTRACT

Many social media sites allow users to upload text, images, and videos (collectively referred to as *anchor post*) for public consumption. These posts may attract hundreds of comments from many social users leading to social conversations (*i.e.*, discussions). Tools that can facilitate user-friendly and effective understanding and analysis of large volumes of comments associated with anchor posts can be of great benefit to individuals and organizations. In this demonstration, we present a novel end-to-end visualization system called KANDINSKY to support *multi-faceted* visualization of social discussions associated with an anchor post. In KANDINSKY, the social discussion landscape is visualized using a collection of colorful *circles* and *concentric circles*, which are inspired from the famous abstract arts called “*Squares with Concentric Circles*” and “*Several Circles*” by Russian painter Wassily Kandinsky (1866-1944). Intuitively, a circle and a concentric circle represent a social comment and a collection of comments in a discussion thread, respectively. We discuss various innovative features of KANDINSKY and demonstrate its effectiveness.

## ACM Reference Format:

Christina Lui, Sourav S Bhowmick, and Adam Jatowt. 2019. KANDINSKY: Abstract Art-Inspired Visualization of Social Discussions. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19)*, July 21–25, 2019, Paris, France. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3331184.3331411>

## 1 INTRODUCTION

Social media sites (*e.g.*, *Facebook*, *YouTube*) allow an individual or an organization to upload post, images, or videos (we referred to it collectively as *anchor post*) for public consumption (*e.g.*, <https://www.youtube.com/watch?v=A838gS8nwas>). Such anchor post may attract hundreds of comments from many commentators (*i.e.*, users) leading to social conversations (*i.e.*, discussions) among them. Since it is time-consuming and cumbersome to comprehend a large number of comments in a social discussion by simply browsing them, tools that can facilitate user-friendly and effective comprehension and analysis of these discussions can be of great benefit to individuals and organizations alike. For instance, government

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*SIGIR '19*, July 21–25, 2019, Paris, France

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6172-9/19/07...\$15.00

<https://doi.org/10.1145/3331184.3331411>



Figure 1: *Squares with Concentric Circles* (1913) and *Several Circles* (1926) by W. Kandinsky [1, 2].

agencies can gain greater insights on issues related to their policies from key discussions associated with their anchor posts.

Unfortunately, majority of the social media sites provide primitive support for such tools. For example, *YouTube* allows one to sort discussions related to an anchor post by “*top comments*” or “*newest first*”. These tools do not allow us to visualize a bird’s-eye view of the conversation landscape, highlighting discussion threads that have garnered considerable attention and interactions between various comments and users. Furthermore, they do not allow visualization of evolution of discussions over time, keyword search on the discussion threads, or visualizing comments that are topically similar in content to a comment-of-interest. Clearly, an easy-to-use, effective visualization system that can support these features will enable us to better comprehend the social conversation landscape associated with an anchor post.

At first glance, it may seem that we can visualize a social conversation using a network [6] where nodes represent comments and edges represent interactions between them (*e.g.*, replies, mentions, retweets). Although such network view is reasonable to visualize social conversation, it becomes cognitively challenging to an end user when the network size becomes large [5]. Hence, traditional network view becomes a stumbling block for comprehending interactions in a large social conversation involving many comments.

In this demonstration, we depart from the traditional network view by presenting a novel end-to-end social conversation visualization system called KANDINSKY that is inspired by abstract art. Specifically, our visualization scheme is based on a collection of colorful *circles* and *concentric circles*, which are based on the famous abstract arts entitled “*Squares with Concentric Circles*” and “*Several Circles*” (Figure 1) by Russian painter and art theorist Wassily Kandinsky (1866-1944). In KANDINSKY, the social conversation landscape with respect to an anchor post is represented as a collection of circles and concentric circles. Specifically, each social comment is represented by a colored circle. Replies to a comment are visualized

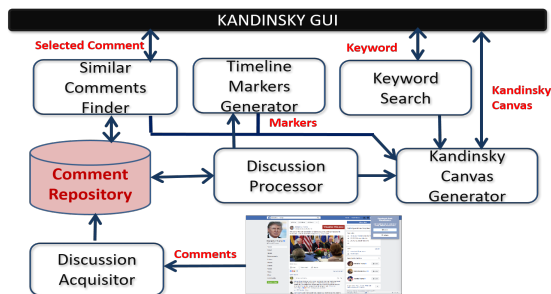


Figure 2: Architecture of KANDINSKY.

as concentric circles with the *original comment* (i.e., comment which directly responds to an anchor post) being the innermost circle. The colors are used as identifiers of the commentators. Hence, more users are involved in a discussion, the larger and more colorful the concentric circles are. To the best of our knowledge, KANDINSKY is the first system that takes such a concentric-circles paradigm for visualizing online social conversations.

## 2 SYSTEM OVERVIEW

Figure 2 shows the architecture of KANDINSKY, which mainly consists of the following modules.

**The Discussion Acquisitor module.** Given a social post/video (referred to as an *anchor post* in the sequel) selected by a user, this module retrieves all comments associated with the post via the social media data API. Specifically, this includes the textual content of a comment, the author information (identifier and name), the date/time of comment, number of likes it has garnered, and the *reply id* in which the comment is replied to.

**The KANDINSKY GUI module.** Figure 3 depicts the GUI of KANDINSKY. It consists of four main panels. Panel 1 visualizes the discussions associated with an anchor post as a collection of colorful circles and concentric circles on a black background. We refer to this panel as the *Kandinsky Canvas*. Such colorful objects on a contrasting background may potentially summon visual attention of end users [9]. Panel 2 shows the *timeline* component of the GUI. It is visualized as a slider and is the starting point where users can kick off exploration of the social conversation landscape. The *span* of the timeline is defined as the time when the first comment was posted to the time of the last comment prior to the retrieval of the comments. The main function of the timeline is to enable users to visualize conversation at a specific time period. Specifically, one can drag on the track of the slider to a desired time for viewing the conversation that occurred during that time period. She can also click on the player icon to auto play the evolution of the social conversation over time. *Timeline markers* (markers for brevity) are introduced on the timeline to indicate significant conversation patterns, which are shown as ticks in different colors. When a user drags the slider to a marker, she can visualize interesting changes to the conversation at that timepoint. These markers are generated by the *Timeline Markers Generator* module.

Panel 3 displays the comments in a selected discussion thread. It is invoked on demand when a user selects a circle/concentric circle in the *Kandinsky Canvas*. This panel can also be used to view similar comments to a user-selected comment by invoking the *Similar Comments Finder* module. Lastly, Panel 4 allows a user



Figure 3: GUI of KANDINSKY.

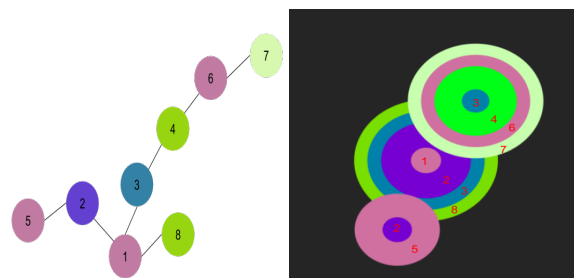


Figure 4: (a) A conversation network (left). (b) Concentric circles (right).

to perform keyword search on the comments associated with an anchor post by invoking the *Keyword Search* module.

**The Discussion Processor module.** The goal of this module is to represent the structural and temporal information associated with discussions to aid effective visualization. Specifically, it aims to extract (a) the chronology of social comments that captures the temporal information of conversations and (b) the connectivity between the comments. The former can be obtained from either the social media platform APIs (e.g., *Facebook*, *YouTube*) or sorting the retrieved raw data (e.g., *Twitter*). Each comment is annotated with the number of likes (if any) it has garnered and user name/id.

Online discussions typically adopt a hierarchical structure with the *original comment* as the root. We refer to it as *conversation network*. Since such hierarchy may result in a large number of levels, several social media platforms constrain the height of the hierarchy using name tagging or hashtag. For instance, each reply to a comment may be referred by the tagged name at the beginning of the reply. Hence, the original structure of the hierarchy is restored in KANDINSKY by connecting the parent node and its children identified by name tagging. This module assumes that a reply by profile tagging is only within a thread, i.e., inter-thread profile reference (i.e., mention) is not considered as reply. Mentions are separately processed to form connections between relevant comments in the network. Figure 4(a) is an example of a conversation network rooted at Comment 1 (i.e., original comment) where ids to comments are assigned according to their creation time and different color codes are assigned to different users.

**The Similar Comments Finder module.** The goal of this module is to find comments that are topically similar to a given comment

in the conversation landscape. To this end, it adopts the Bags-of-Words model to represent each comment. Since social comments can be noisy, they are first *cleaned* and *processed* to extract key phrases in the content. Specifically, hyperlinks, repetitive letters, and non-ASCII characters (e.g., emojis, non-English characters), profanities, and internet slangs are removed. English enclitic (i.e., abbreviated forms) are converted to their full form and noun phrase chunking is performed for keyword/keyphrase extraction. Then, topics associated with each comment are modelled using LDA with Gibbs Sampling method<sup>1</sup>. The number of topics  $K$  is set to 10 as it gives a good balance between processing time and log likelihood value. The hyperparameters  $\alpha$  and  $\beta$  are set to 0.1 and 0.01, respectively, as they give good distribution of words to describe a topic in reasonable time. Hence, after this step, each comment is associated with a topic distribution.

Given a user-selected comment  $c$ , this module computes the cosine similarity between the topic distributions of  $c$  and the comments in the *Comment Repository* and selects those with similarity score greater than the *similarity threshold*  $\delta$  ( $\delta = 0.5$  as default).

**The Keyword Search module.** This module implements keyword-based search on comments associated with an anchor post by utilizing inverted index.

**The Timeline Markers module.** The goal of the *Timeline Markers Generator* module is to generate the *timeline markers* automatically from the underlying conversation data. *Timeline markers* are indicators shown on the timeline (Panel 2 in Figure 3) to mark the time when some interesting pattern w.r.t the social conversation is detected (e.g., a surge of comments). These markers are generated by first transforming the user interaction data into feature vectors by partitioning the timestamps with a time window  $W$  and then clustering them with  $k$ -means clustering using Lloyd’s method [7].

**The Kandinsky Canvas Generator module.** The *Kandinsky Canvas* is generated to visualize various aspects of a social conversation landscape. Each comment is represented by a colored circle. The radius of a circle is influenced by normalized value of the *like* count of the comment, which is a measure of its popularity. Color codes are generated randomly such that comments posted by the same author share the same color. Each comment is also tagged with a unique id that is assigned to each author’s name. Hence, if an anchor post only contains original comments with no replies to them then the *Kandinsky Canvas* will contain a collection of circles with varying color and size.

In reality, an original comment may garner responses from different social users resulting in a conversation network with the original comment as the root. In KANDINSKY, replies to a comment are visualized as concentric circles. The more colors a concentric circles contain, the more social users participated in the discussion. Circles in a concentric circle are added according to the chronological order of the associated comments (ties are broken arbitrarily). The innermost circle of a concentric circle is called a *pivot*. Intuitively, KANDINSKY displays two types of concentric circles, namely *nucleus* and *peripheral*. Concentric circles resulted from an original comment and its replies is referred to as *nucleus*. In other words, in a conversation network, the root and its children form a nucleus

with the root being the pivot. On the other hand, replies at depth equal to or greater than 2 in a conversation network (e.g., replies to a reply) form separate concentric circles, referred to as *peripherals*. Peripherals are positioned on the periphery of a nucleus. For instance, consider Figure 4(b), which visualizes the conversation network in Figure 4(a) utilizing a group of concentric circles. The color codes of the circles follow that of the comments. The nucleus is created using Comments 1 (pivot), 2, 3, and 8. Note that 1 is the original comment whereas 2, 3, and 8 are replies to it. On the other hand, since the depths of Comments 4, 5, 6, and 7 in the network are greater or equal to 2, two peripherals are generated with Comments 2 and 3 being pivots, respectively. Observe that a nucleus represents an original comment that responds to an anchor post (i.e., an initiator of discussions) whereas a peripheral represents responses and discussions triggered by an original comment. These collection of circles and concentric circles are positioned on the *Kandinsky Canvas* by exploiting force-directed graph drawing.

Observe that the aforementioned visualization scheme on *Kandinsky Canvas* does not display the textual contents of comments. A user may wish to view specific comments related to a discussion thread. However, it can be cognitively challenging to visualize all comments on a GUI as an anchor post may garner hundreds of comments. Hence, we adopt the “details on demand” interface design strategy for visualizing textual content by allowing a user to select a part of a discussion in the *Kandinsky Canvas* that she may be interested in. Specifically, when a user selects a circle or a concentric circle, the corresponding discussion thread will be revealed in Panel 2, where each comment text is assigned a color bar that corresponds to its color in the circle (or concentric circles) and its thickness is proportional to its normalized like count. Moreover, topic-specific keywords are highlighted with yellow color in each comment. When a user selects a particular comment in Panel 2, a list of similar comments from other discussion threads will be shown by invoking the *Similar Comments Finder* module. The corresponding circles containing these similar comments are highlighted with yellow bounding boxes. Note that in order to avoid retrieving too many comments and cluttering the GUI, we retrieve top-10 comments that satisfy the similarity threshold  $\delta$ .

A user may enter a keyword query in Panel 4 and comments containing it will be highlighted in white bounding box for easy visualization. One may click on any of these highlighted comments to view its textual content containing the keyword.

To visualize the differences of a discussion landscape at two different timestamps  $t_1$  and  $t_2$ , this module allows a user to first load the *Kandinsky Canvas* at  $t_1$  and then at  $t_2$ . Opacity of the colors of comments that exist in both  $t_1$  and  $t_2$  are subsided and only new comments that occurred between these two timepoints are shown in brighter colors.

### 3 RELATED SYSTEMS AND NOVELTY

Several works in the literature have focused on visualizing topics [8] and networks [6]. More germane to this demo are efforts on visualizing social conversations [3, 4, 10]. Björn et. al. [4] used *Squarified Treemap Layout* to visualize threaded discussion forums on devices with small screen size. Intuitively, each tile in the *Treemap* corresponds to a node in the network. The position of the tile indicates the importance of the thread, where the most discussed thread is

<sup>1</sup>The KANDINSKY framework is orthogonal to the choice of text cleaning and topic modeling techniques. Hence, any superior text cleaning or topic modelling technique designed for social comments can be plugged into the framework.

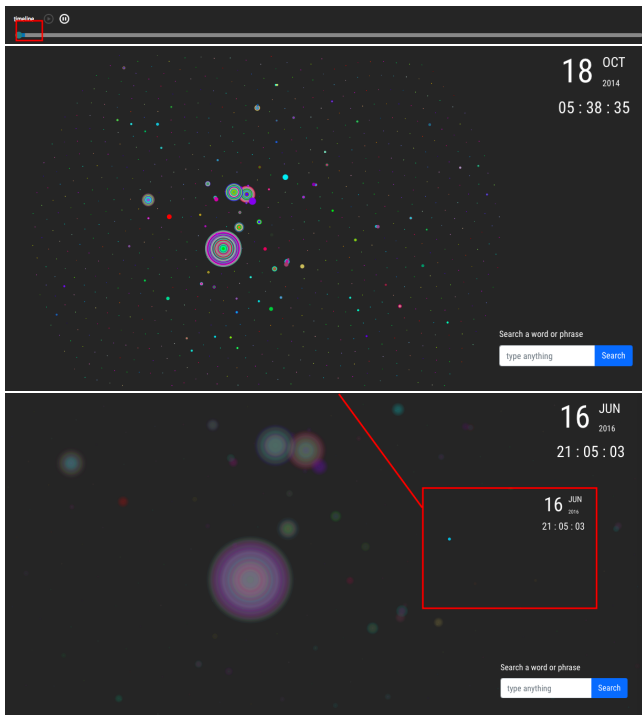


Figure 5: *Kandinsky Canvas* of Alex Jones's post and evolution of the conversation.

placed in the upper left corner and those that are less active are in the right bottom. *Conversation landscape* [3] is a 2D visualization of a chat group where the x-axis and y-axis represent participants of a conversation and timeline of postings by participants, respectively. Each post is represented by a horizontal bar and the length of a bar indicates the post length. *PeopleGarden* [10] portrays users' interactions on social environment over time as a garden of flowers where each flower represents a user and each petal is a post by him/her. A flower "blossoms" as a user becomes more engaged in a discussion. These approaches, however, do not leverage on circles and concentric circles for visualization. Moreover, they do not effectively capture the contextual details and temporal differences of a discussion. KANDINSKY not only supports temporal view but also enables retrieval of contextual details by supporting features such as finding similar comments and keyword search.

#### 4 DEMONSTRATION OVERVIEW

KANDINSKY is implemented using HTML/CSS and Javascript and is running as a Chrome extension. Our demonstration will make available to the audience a few real anchor posts with different number of comments in various social platforms. Users can also select their own anchor posts. The key objective of the demonstration is to enable the audience to interactively experience various features of KANDINSKY. We illustrate them with an example. A short video of KANDINSKY is available at <https://youtu.be/eIDpB51m3Tc>.

On 13th October 2014, controversial radio host *Alex Jones* published a post (*i.e.*, anchor post) on *Facebook*, which discussed whether Ebola was invented to depopulate, after American singer *Chris Brown* claimed Ebola is a "form of population control". The Facebook post garnered about 1,200 comments. Figure 5 (top) depicts

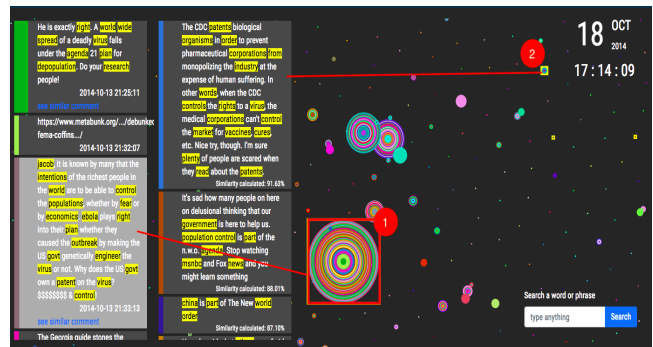


Figure 6: Similar comments to a selected comment (in grey) are shown in the middle panel. The locations of these comments are shown with yellow bounding boxes in the canvas.

the *Kandinsky Canvas* of this post. Observe that a single-color circle indicates a comment. The larger a circle, the more "likes" the corresponding comment has garnered. Each concentric circle has several colors as several users responded to the original comment.

**Evolution of discussion threads.** Figure 5 shows the evolution of the discussion. The marker highlighted in red shows that major discussions occurred within a week after the post. The two *Kandinsky Canvases* at two different timepoints (18th Oct, 2014 and 16th June, 2016) look almost the same – only a comment was posted (highlighted by the red rectangular box) during this time.

**Similar comments.** A key feature of KANDINSKY is to enable a user to quickly identify and browse discussion threads that have garnered considerable comments (*e.g.*, bounded by the red box in Figure 6). In particular, the concentric-circles view allows a user to quickly zero in to such popular discussion threads. While browsing *Discussion 1*, one may observe that one of the comments mentioned about the U.S. Government's ownership of patent of Ebola virus. The user finds this intriguing and would like to find out if there is any other comment that is similar to this. The most similar comment found is related to patent ownership by CDC (marked as 2 in Figure 6), which is discussed in a separate discussion thread.

**Keyword search.** KANDINSKY supports keyword search on the discussion threads. For example, is it a conspiracy that the U.S. Government is using Ebola as a population control? By searching for "conspiracy", we can visualize that there are several threads (highlighted with white bounding box) involved in this discussion.

#### REFERENCES

- [1] W. Kandinsky. Squares with concentric circles. <http://www.kandinskypaintings.org/squares-with-concentric-circles/>.
- [2] W. Kandinsky. Severak circles. [https://en.wikipedia.org/wiki/File:Vassily\\_Kandinsky,\\_1926\\_-\\_Several\\_Circles,\\_Gugg\\_0910\\_25.jpg](https://en.wikipedia.org/wiki/File:Vassily_Kandinsky,_1926_-_Several_Circles,_Gugg_0910_25.jpg).
- [3] J. Donath, K. Karahalios, F. B. Viégas. Visualizing Conversation. In *HICSS*, 1999.
- [4] B. Engdahl, M. Köksal, G. Marsden. Using Treemaps to Visualize Threaded Discussion Forums on PDAs. In *CHI*, 2005.
- [5] W. Huang, P. Eades, S.H. Hong. Measuring Effectiveness of Graph Visualizations: A Cognitive Load Perspective. *Inf. Vis.*, 8(3), 2009.
- [6] C. D. G. Linhares, B. A. N. Travençolo, et al. DyNetVis: A System for Visualization of Dynamic Networks. In *ACM SAC*, 2017.
- [7] S. Lloyd. Least Squares Quantization in PCM. *IEEE Trans. Inf. Theor.*, 28(2), 2006.
- [8] T. Mitja, M. Grobelnik. Visualization of Online Discussion Forums. In *Proceedings of the First Workshop on Applications of Pattern Analysis*, 2010.
- [9] M. Turatto, G. Galfano. Color, Form and Luminance Capture Attention in Visual Search. *Vision Research*, 40(13), 2000.
- [10] R. Xiong, J. Donath. PeopleGarden: Creating Data Portraits for Users. In *UIST*, 1999.