From community search to community understanding: A multimodal community query engine

Zhao LI

Pengcheng ZOU

Xia CHEN

Shichang HU

Peng ZHANG

See next page for additional authors

Author
Zhao LI, Pengcheng ZOU, Xia CHEN, Shichang HU, Peng ZHANG, Yumou ZHOU, Bingsheng HE, Yuchen LI, and Xing TANG

From Community Search to Community Understanding: A Multimodal Community Query Engine

Zhao Li[1*], Pengcheng Zou[1], Xia Chen[1], Shichang Hu[1], Peng Zhang[2], Yumou Zhang[1], Bingsheng He[3], Yuchen Li[4], Xing Tang[1]
[1]Alibaba Group, Hangzhou, China [2]Guangzhou University, China
[3]National University of Singapore, Singapore [4]Singapore Management University, Singapore
{lizhao.lz,xuanwei.zpc,xia.cx,shichang.hsc,yumou.zhangym,pingchou.pc}@alibaba-inc.com
p.zhang@gzhu.edu.cn,hebs@comp.nus.edu.sg,yuchenli@smu.edu.sg

ABSTRACT

In this demo, we present an online multi-modal community query engine (MQE[1]) on Alibaba’s billion-scale heterogeneous network. MQE has two distinct features in comparison with existing community query engines. Firstly, MQE supports multimodal community search on heterogeneous graphs with keyword and image queries. Secondly, to facilitate community understanding in real business scenarios, MQE generates natural language descriptions for the retrieved community in combination with other useful demographic information. The distinct features of MQE benefit many downstream applications in Alibaba’s e-commerce platform like recommendation. Our experiments confirm the effectiveness and efficiency of MQE on graphs with billions of edges.

CCS CONCEPTS

• **Information systems → Electronic commerce**; **Multimedia and multimodal retrieval**; • **Computing methodologies → Natural language generation**.

KEYWORDS

Community Search, Community Understanding, Multimodal Search, Text Generation

ACM Reference Format:
Zhao Li[1*], Pengcheng Zou[1], Xia Chen[1], Shichang Hu[1], Peng Zhang[2], Yumou Zhang[1], Bingsheng He[3], Yuchen Li[4], Xing Tang[1]. 2021. From Community Search to Community Understanding: A Multimodal Community Query Engine. In *Proceedings of the 30th ACM Int’l Conf. on Information and Knowledge Management (CIKM ’21), November 1–5, 2021, Virtual Event, Australia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3459637.3481973

1 INTRODUCTION

Complex information on e-commerce platforms such as Alibaba and Amazon can be naturally represented as a large network, where buyers, items and sellers are denoted as nodes, and links of clicks, purchases, and collections of items are denoted as edges. On such platforms, communities are groups of buyers, items and sellers that are densely connected, and it is necessary to effectively organize the nodes into communities for downstream tasks. For example, in a fast moving market, it is often the case that a seller wants to rapidly identify the target groups of users for different groups of items when designing product promotion strategies, and a buyer wants to find the groups of buyers sharing similar preferences when making purchasing decisions. Although there has been a large amount of effort in solving the community search [4] problem, three major **limitations** still remain unsolved when supporting real-world business applications.

The first limitation is to support **multimodal queries**. Existing community search methods can be generally grouped into the following types: k-core algorithms, k-truss algorithms , k-clique algorithms, and local modularity algorithms, etc. [4]. These algorithms can only support queries of vertices. However, e-commerce platforms inherently have rich multimodal data including images and texts. Thus, the community search engine should enable multimodal query input such as keyword and image queries. The second limitation is to support **heterogeneous networks** [5, 6]. Most existing community search methods only consider labeled but homogeneous networks. In a heterogeneous network, there may be multiple node types and edge types, that should be treated discriminatively. For example, for the k-core-based community search algorithms [4], the number of neighbors should be judged separately by different neighbor types. Although there have been a few studies on community search in heterogeneous networks [5, 6], all of them cannot solve multimodal queries.

The third limitation is to support **community understanding**. One straightforward approach is to display multi-dimensional statistical information calculated from community demographic. However, the complex statistics may easily overwhelm users in a real environment, especially those from a non-technical background, e.g., crowd-sourced workers recruited for community labeling on recommendation tasks. It thus calls for an automatic generation of concise natural language descriptions on communities to smooth the process of user understanding. In particular, a textual description of community facilitates users to in-depth understanding. In the literature of text generation, the sequence to sequence model provides the basic generation framework [8]. Recently, Transformer models such as BERT [2] rapidly become popular. Moreover, knowledge based transformers are widely used to strengthen text generation by including external knowledge.

[1]An online video of MQE can be found at http://xuanwei.cn-hangzhou.oss.aliyun-inc.com/cikm2021/MQE_demo.mp4

CIKM ’21, November 1–5, 2021, Virtual Event, Australia.

ACM ISBN 978-1-4503-8446-9/21/11…$15.00
https://doi.org/10.1145/3459637.3481973

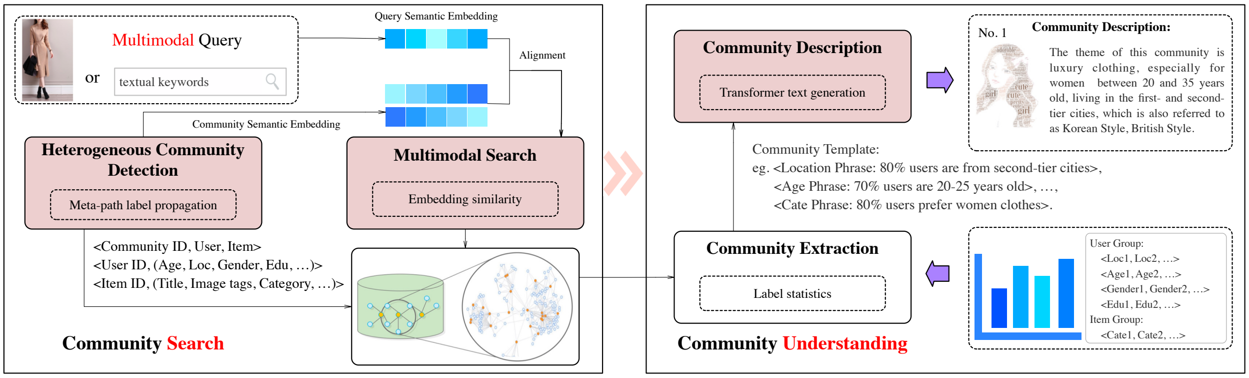

Figure 1: The architecture of Multi-modal Query Engine (MQE). MQE enables multimodal community search on a heterogeneous network, and then uses text generation to help understanding the returned communities.

In light of the above limitations, we propose a new graph querying system called *MQE*. MQE supports multimodal queries including keyword and image queries. Moreover, MQE enables an end-to-end data science pipeline from search, visualization to understanding of the communities in large-scale heterogeneous networks. The technical contributions are summarized as follows:

Multimodal community search on heterogeneous networks. MQE enables online heterogeneous community search with multimodal queries. A novel meta-path based label propagation algorithm is proposed to find the candidate communities offline. The candidate communities and queries (e.g., keywords and images) are transformed into embeddings using BERT [2], and the online search is efficiently processed in the embedding space. This design allows real-time performance of community search on large e-commerce networks with *hundreds of millions of* items and users and *billions of* edges.

Community understanding. MQE generates concise descriptions of the returned communities based on the multimodal queries. We devise a Transformer-based model [1] to generate natural language descriptions of the returned communities. To the best of our knowledge, this is the first systematic study on community understanding with natural language descriptions. Our A/B test shows that MQE boosts the Click Through Rate by 4.26% and thus provides better overall user experience.

In the demo, audiences are able to specify their own queries and experience with these two techniques from community search and understanding.

2 MULTIMODAL QUERY FRAMEWORK

Figure 1 shows the architecture of MQE in Alibaba. The system mainly consists of two modules: community search and community understanding. In the community search module, MQE enables online multimodel community search on heterogeneous networks by first detecting all the communities based on a meta-path label propagation in an offline manner, and then querying these candidate communities with either keywords or images in an online manner. In the community understanding module, the returned communities are visualized and textually described.

2.1 Multimodal Community Search

In the multimodal scenario, the traditional community search algorithms cannot be performed without query nodes. To search communities with multimodel queries (e.g., texts and images) online, MQE introduces an novel online multimodal community search framework, by employing offline community detection and online vector learning of communities and multimodal queries.

Given a heterogeneous network $G = \{V, E, Z, R\}$ where V denotes a set of nodes, E denotes a set of edges, Z denotes node types, and R represents edge types. We propose a new meta-path based label propagation algorithm to find the candidate communities offline. Specifically, we combine meta-path[3] and SLPA [9] to efficiently discover candidate communities from the heterogeneous network, where SLPA is a highly efficient label propagation algorithm for detecting overlapping communities.

A meta-path $\mathcal{P}$ can be defined as $V_1 \xrightarrow{R_1} V_2 \xrightarrow{R_2} ...V_z \xrightarrow{R_z} V_{z+1} ... \xrightarrow{R_{l-1}} V_l$, where V_z is the set of nodes of type z, and R_z is the relation of type z. The transition probability of labels on the meta-path $\mathcal{P}$ is defined as follows:

$$p(v|v_z, \mathcal{P}) = \begin{cases} 1, & (v, v_z) \in E, v \in V_{z+1} \\ 0, & (v, v_z) \in E, v \notin V_{z+1}, \ or \ (v, v_z) \notin E \end{cases} \quad (1)$$

where $v_z \in V_z$ denotes a node of type z. Meta-paths are symmetric[3], i.e., *User-Item-User* and *User-Item-Seller-Item-User*. The meta-path and the transition probability can be adjusted for different applications. Figure 2 shows the label propagation with respect to a specific path *User-Item-Seller* of the meta-path *User-Item-Seller-Item-User*, where the labels are first propagated from users to items and then from items to sellers. Specifically, at the first step users send their labels to neighbouring items. Then, the items update their labels according to the labels propagated from their neighbouring users using the transition probability defined in Eq.(1). At the last step, the labels of items are propagated to the their neighbouring sellers.

After finding all candidate communities discovered offline, a multimodal search is designed for online queries with either keywords or images. First, an image query will be first converted into a textual vector based on the multi-label learning model trained on the tuples of "(item images, item tags)". Then, the image vector and the input keywords are processed by a BERT model [2] to generate the query

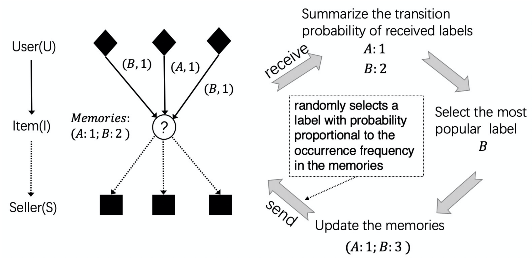

Figure 2: An illustration of heterogeneous label propagation by a path *User-Item-Seller.*

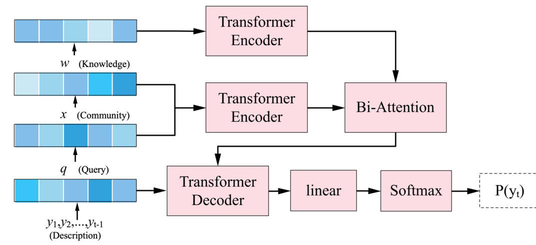

Figure 3: A Transformer model for generating textual descriptions for communities.

embedding vector. For each candidate community generated by the community detection model, a community embedding vector is also generated by the BERT model based on the item titles, item images and image tags in the community. Finally, the community search module returns relevant communities based on the similarity between the embedding spaces of multimodal queries and candidate communities.

To find the most relevant communities, a data structure of the candidate community vectors $\boldsymbol{e}^C$ is created as an indexed. When a new query k is submitted and the query is represented by a vector e_k^Q, the engine computes the most relevant community $i = \arg\min_i \ \|e_i^C - e_k^Q\|$, where $\|\cdot\|$ is an Euclidean distance in the embedding spaces. By computing the argmin results, we can obtain the most relevant community Q_i for the query q_k.

2.2 Community Understanding

The returned communities are a set of nodes and edges, which are very difficult to understand by non-technical users in the downstream applications. To facilitate the understanding of the communities, a concise summarization of each returned community is required. Although a large number of text generation models [2] have been proposed, description generation for network communities has not been addressed yet. To generate meaningful and personalized descriptions for communities, we devise a Transformer-based model [1] which encodes the query information, the statistical descriptive phrases extracted from the communities and an external knowledge base, as shown in Figure 3.

Community Extraction This component extracts statistical information of the returned communities which include user groups (e.g., user's location, age, gender and education) and item groups (e.g, item's category). The statistics is converted into textual phrases using templates. For example, a location description can be "80% of users are from second-tier cities", where 80% and the second-tier are the statistical information. Similarly, an example description on age can be "70% of users are 20-25 years old", and an example description on item can be "80% of users prefer women's dresses".

Community Description We introduce a Transformer model [1] to generate textual descriptions for the returned communities. Formally, we aim to generate a sequence of descriptive words $Y = (y_1, y_2, ..., y_m)$ based on each returned community X represented as a sequence of words $(x_1, x_2, ..., x_n)$ and the query q represented as a sequence of words $(q_1, q_2, ..., q_n)$. The description also borrows the strength of an external Knowledge base W (such as CN-DBpedia and Wikipedia). Therefore, the inputs consist of three parts: the returned community x, a textual query q and a knowledge base W. The training objective function is based on the maximum likelihood estimation of $P(y|x, q, W) = \prod_{t=1}^{m} P(y_t|y_1, y_2, ..., y_{t-1}, x, q|W)$, where the objective is to generate a sequence that can approximate the target Y based on the input sequences.

3 DEMONSTRATION OVERVIEW

We implement MQE by deploying it on the Alibaba distributed graph platform (Alibaba ODPS GRAPH) and develop the graphical user interface. We demonstrate the results with respect to graphical user interface and visualization, system performance analysis and community understanding and application. The new engine is tested on an Alibaba transaction network with approximately 501 millions of nodes and 8.58 billions of edges.

Audience Interaction. We aim to provide rich experience for conference audiences to interact with our demo system. Audiences can easily issue queries by giving keyword queries or uploading any photos to find relevant communities during the conference event.

As shown in Figure 4, given a query "Spring outfit", the most relevant communities will be displayed. In this demo, the item size and user size of discovered communities generally ranges from dozens to thousands. Clearly visualize the whole community is difficult. Thus, random sampling of nodes is used. In addition, we have hidden the nodes of sellers in the visualization, because it is not needed for practical applications and makes the visual network more cumbersome.

For audiences to quickly grasp the unique characteristics of a community, MQE automatically generates a concise community description as well as a cartoon character depicting the profile of users in the community. Further, audiences can interact with the interface by exploring the community network structure as well as community statistical information.

System performance analysis. MQE discovers hundreds of thousands of offline communities in the billion-scale network of Alibaba. We also extract an example of 3,395 items(I), 1,056 users(U), 1,367 sellers(S) and 10,186 edges as the testbed for evaluating the performance of community detection. For fair comparison, we also use the popular DBLP dataset which contains 1,000 papers(P), 1,587 authors(A), 2,096 terms(T) and 10,066 edges for testing.

We evaluate MQE (SLPA + metapath) for heterogeneous community detection. The SLPA, COPRA and WLPA [7, 9]) algorithms

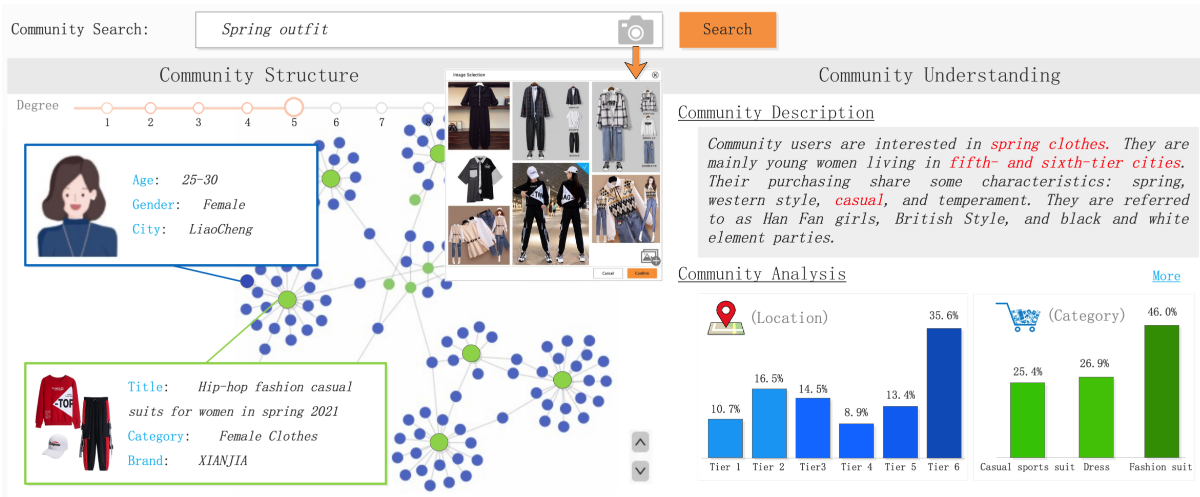

Figure 4: An example of the query results. Given a query keyword "Spring outfit" or a query image of "Spring outfit", a set of relevant communities will be returned (due to space, we only show the top-1 most relevant community). An overview of the network structure and attributes of the nodes are shown on the left, and the textual description and statistical analysis of the community are shown on the right.

are used as the baseline. Since original COPRA and WLPA algorithms cannot handle heterogeneous networks, we combine the meta-path with COPRA and WLPA as the compared methods. The performance is evaluated with the heterogeneous overlapping modularity, which is defined by combining the overlapping modularity [9] of homogeneous and bipartite networks. The meta-paths used in AE-s1 and DBLP are *UISIU* and *APTPA*, respectively. The reason is that the meta-paths can go through all the types of nodes with the shortest path.

From the results in Table 1, MQE obtains better results than SLPA, COPRA and WLPA. WLPA obtains the same results with COPRA. The reason is WLPA is an extended version of COPRA and the optimization doesn't work in heterogeneous networks. Moreover, we run the algorithm with different network sizes extracted from the Alibaba platform to test the runtime. Table 2 shows the size of used networks and the runtime. MQE can efficiently update the community detection results on heterogeneous networks with hundreds of millions of nodes and billions of edges per hour.

When an image or textual keyword is given, MQE returns the communities in 500 milliseconds, which guarantees the efficiency of online search.

Community understanding and application. We report the generated community description and send it to 50 human experts for evaluation. These experts are interested in the description which can help better understand their customers and polish their promotion strategy, as well as increase revenue for Alibaba. During the evaluation, a number of generated communities and the corresponding statistical analysis and descriptions are reported on a questionnaire. They complete the questionnaire based on whether they are satisfied with the returned communities and descriptions. The feedback shows that **85%** of the experts are satisfied with the descriptions of the communities. For example, when input a query keyword "Spring outfit", the experts can efficiently obtain the girls who love Korean style clothing. To verify the effectiveness of MQE, we conduct an online A/B testing to deliver a set of given items to two groups of users, i.e., a controlled group and an experimental group. In the controlled group, the users are selected by the human experts, while the users in the experimental group are generated by MQE. The results of the A/B testing demonstrate that MQE can boost the Click Through Rate (CTR) by 4.26%.

Table 1: Comparison *w.r.t* community modularity.

Dataset	SLPA	COPRA	WLPA	**MQE**
AE-s1	0.2279	0.2379	0.2379	**0.4050**
DBLP	0.3552	0.3659	0.3659	**0.5333**

Table 2: Community Detection Time (Offline).

dataset	nodes(millions)	edges(millions)	time(seconds)
AE1	4.12	48.11	204
AE2	65.25	402.62	794
AE3	125.84	939.79	1800
AE4	338.73	3763.04	4878
AE5	501.01	8586.36	15127

4 CONCLUSION

We demonstrate a multimodal query engine (MQE) for community search and understanding on Alibaba's billion-scale heterogeneous network. As shown in the demo and video, MQE is efficient and effective. The community understanding is helpful to boost the CTR of item recommendation in e-commerce platform.

REFERENCES

[1] Qibin Chen, Junyang Lin, Yichang Zhang, Hongxia Yang, Jingren Zhou, and Jie Tang. 2019. Towards knowledge-based personalized product description generation in e-commerce. In *SIGKDD*. 3040–3050.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL*. 4171–4186.

[3] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *SIGKDD*. 135–144.

[4] Yixiang Fang, Xin Huang, Lu Qin, Ying Zhang, Wenjie Zhang, Reynold Cheng, and Xuemin Lin. 2020. A survey of community search over big graphs. *VLDB* 29, 1 (2020), 353–392.

[5] Yixiang Fang, Yixing Yang, Wenjie Zhang, Xuemin Lin, and Xin Cao. 2020. Effective and efficient community search over large heterogeneous information networks. *VLDB* 13, 6 (2020), 854–867.

[6] Xun Jian, Yue Wang, and Lei Chen. 2020. Effective and efficient relational community detection and search in large dynamic heterogeneous information networks. *Proceedings of the VLDB Endowment* 13, 10 (2020), 1723–1736.

[7] Meilian Lu, Zhenglin Zhang, Zhihe Qu, and Yu Kang. 2018. LPANNI: Overlapping community detection using label propagation in large-scale complex networks. *TKDE* 31, 9 (2018), 1736–1749.

[8] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to Sequence Learning with Neural Networks. *NeurIPS* 27 (2014), 3104–3112.

[9] Jierui Xie, Stephen Kelley, and Boleslaw K Szymanski. 2013. Overlapping community detection in networks: The state-of-the-art and comparative study. *CSUR* 45, 4 (2013), 1–35.