QeNoBi: A System for QuErying and miNing BehavIoral Patterns Authors' Copy

Abdelouahab Chibah, Sihem Amer-Yahia, Laure Berti-Equille

QeNoBi: A System for QuErying and miNing BehavIoral Patterns Authors' Copy

Abdelouahab Chibah
CNRS, Univ. Grenoble Alpes
Saint Martin D'Hères, France
abdelouahab.chibah@univ-grenoble-alpes.fr

Sihem Amer-Yahia
CNRS, Univ. Grenoble Alpes
Saint Martin D'Hères, France
sihem.amer-yahia@univ-grenoble-alpes.fr

Laure Berti-Equille
IRD, ESPACE-DEV
Montpellier, France
laure.berti@ird.fr

***Abstract*—We demonstrate QeNoBi, a system for mining and querying customer behavioral patterns. QeNoBi combines an interactive visual interface, on-demand mining, and efficient top-k processing, to provide the exploration of customer behavior over time. QeNoBi relies on two distinct data models: a customer-centric graph that represents customers with similar purchasing behaviors and is annotated with a change algebra to reflect their behavior evolution, and product-centric time series that reflect the evolution of customer purchases over time. Users can query both representations along three dimensions: shape (the sketched trend of the behavior), scope (the set of customers/products of interest), and time granularity. QeNoBi provides a holistic behavior exploration capability by allowing users to seamlessly switch between customer-centric and product-centric views in a coordinated manner, thereby catering to various needs. A demonstration of QeNoBi is available at https://bit.ly/2HlcO3S.**

I. Introduction

As user data is becoming available in large volumes, new opportunities for mining and exploring behavioral patterns arise. While the focus has mostly been on mining customer group evolution in the retail domain [5], [9], [7], [6], there is a need for a flexible tool for the on-demand discovery of behavioral patterns and their exploration over time in e-commerce and beyond. Such functionality benefits a variety of stakeholders. It enables the design of marketing strategies, population studies, and various campaigns. Data scientists, social scientists, Web application designers, marketers, and other domain experts, need a single destination to explore behavioral changes. We will demonstrate QeNoBi, a system that enables the on-demand querying and exploration of behavioral changes by providing a visual interface that lets users combines filters on customers and products, time windows, and types of change, at every stage of the exploration. Although the demo setting is specific to customer/product type of datasets, the application of QeNoBi can be generalized to a variety of user studies based on time series.

Illustrative Example. Consider a product manager interested in the adoption of hand sanitizers, a newly introduced product following the COVID-19 pandemic. The manager first explores a product-centric time series that represents the number of unique customers who bought hand sanitizers over a period of interest. The manager can specify filtering conditions on customer attributes such as age group and location, and query time series to identify early adopters of hand sanitizers. Next, the manager requests to switch to a customer-centric view in the form of a Sankey diagram to observe the evolution of customer groups, analyze their purchase frequency, and quantify the strength of adoption and loyalty of new customers over time. Meanwhile, s/he might request a product-centric view of complementary products (e.g., hand creams, masks, lipsticks) and identify a simultaneous decline or rise in their purchases. S/He may also look for specific trends and shapes in the data, (e.g., a steep rise followed by a shallow decline) for specific products and with different time granularities (e.g., per week, month, quarter). Ultimately, s/he requests another customer-centric view to query other visually prevalent patterns across groups.

Our system. QeNoBi addresses the above needs by providing the following features: (1) A visual interface to generate product-centric time series and customer-centric Sankey diagrams to visualize the behavior of customer groups over time; (2) The ability to seamlessly switch between a product-centric view and a customer-centric view to obtain rich insights; (3) A unified interface to query time series and Sankey diagrams across various modalities (i.e., with sketch drawing and primitive declaration) to leverage both shape search and algebraic primitives that augment behavioral pattern discovery from groups of users; and (4) A coordinated view that automatically updates product-centric time series and customer-centric Sankey diagrams when one or the other is queried.

Positioning. There are many tools for visual time series exploration. TimeSearcher [1] lets users apply constraints on the x and y range values via boxes or query envelopes. Qetch [8] and ShapeSearch [12] enable expressive shape queries in the form of sketches, natural-language, and visual regular expressions. RINSE implements adaptive index-based data series exploration [15]. Data Polygamy enables exploring spatio-temporal datasets [2]. Other visual time series exploration tools such as Metro-Viz [3], ONEX [10] and Steiger et al. [13] enable anomaly detection and clustering. QeNoBi differs from these tools in its ability to combine queries on shapes (both

on time series and Sankey diagrams), customers/products and time simultaneously, and in relying on a top-k processing algorithm to return the k best matching segments of the time series.

II. System Architecture and Technical Details

QeNoBi is designed with a modular architecture that integrates two functionalities that have traditionally been studied separately: *mining* and *querying* customer behaviors. Our system presented in Figure 1 has three main modules: a customer-centric module to mine and query Sankey diagrams, a product-centric module to generate and query product times series, and an integrated visual query interface that allows the specification of shapes, customer/product filters, and time granularity, and enables coordinated views between time series and Sankey diagrams. To enable that, raw customer transaction data are used to generate a customer-centric graph model annotated with our change algebra, and a product-centric time series which reflects aggregated purchases over time.

The **product-centric module** of QeNoBi builds on top of Qetch [8], a system for querying time series, with the ability to detect shapes reflecting interest gain and loss in products for a selected set of customers, over a specified time granularity. It supports visual sketches and a custom similarity metric that is robust to distortions in the query. The scoring function captures how well matching segments in a time series reflect an input sketch (as in Qetch), but also how well it matches product/customer filters, and the specified time granularity. We implemented a Fagin-style top-k algorithm to return the k best segments.

The **customer-centric module** of QeNoBi enables the generation and visualization of a Sankey diagram that represents the evolution of groups of customers over consecutive time periods. The edges of the diagram are labeled with change primitives (defined in Table 1) to represent groups that are stable, grow, merge, split, or perish. Users can query the Sankey diagram by specifying a triple (shape, scope, granularity) where shape reflects a series of changes of interest for a set of customers/products during consecutive time periods of a given time granularity. This work is closely related to customer behavioral analytics and user group analysis, which consist in breaking down customers into similar groups to gain a more focused understanding of their behavior [11]. Similarly to [4], [7], [6], we rely on pattern mining [14] to build our groups in the same time period. The modular design of QeNoBi can support a variety of group mining algorithms. To enable a fine-grained representation of change, we extend existing group evolution algebras to capture change between time periods.

Technical contributions: QeNoBi includes an API that generates Sankey graphs from the output of LCM [14] mining algorithm but other mining algorithms could be used. Sankey graphs are annotated with our change algebra and efficiently queried despite multiple explorations of the graph. Conventional queries are extended with filters on behavioral shapes of interest over time with a unified scoring of the resulting data segments combining shape, scope, and time granularity for product-centric views.

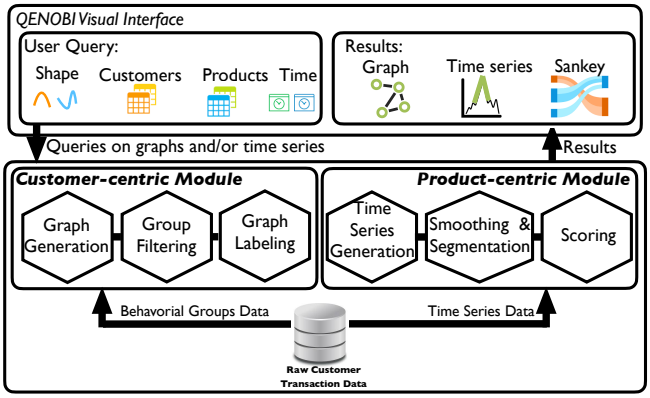

Fig. 1. Architecture of QeNoBi

III. Data Model and QENOBI Views

A. Representing Customer Behaviors

Our data represents customers purchasing products.

A **customer-centric view** of our data is a function $\texttt{CView} : T \times P \times C \rightarrow S \times T$ that admits a time period T, a set of products $P \subseteq \mathcal{P}$, a set of customers $C \subseteq \mathcal{C}$, and produces a sequence of pairs $[(S_1, t_1), \ldots, (S_n, t_n)]$ where each S^i is a set of customer sets. A customer set $s \in S_i$ reflects customers with common demographics and same purchase patterns for some products in P during time period t_i. Each set s is characterized by a vector of values $s.label$ that reflects its common demographics and products. For instance, $s.label = \langle \texttt{Young Adult}, \texttt{South}, \texttt{Hand Sanitizer} \rangle$ characterizes young adults with similar interest for hand sanitizers.

Given two sets S in period t and S' in its following period t', for each group $s \in S$, we characterize its change semantics between t and t' with the primitives in Table 1 and we construct a behavioral graph, noted G, where each node is a set $s \in S_i$ for all $t_i \in T$. There exists an edge between two sets $s \in S$ in period t and $s' \in S'$ in a consecutive period t' for each primitive that takes s and S' as input and returns a non-empty set in S'. Edges are labeled with the primitive used to generate it. Figure 2 shows a Sankey representation of a labeled graph. The graph represents the evolution of a set of young adults ($<$ 35 years old) who purchase hand sanitizers over a period of interest. The interface uses enriched Sankey diagrams to represent behavioral changes and a color coding scheme on edges to reflect different primitives of our algebra. Users can hover over one group and see details in the right panel. They can additionally interact with the graph and specify queries in the form of shape that combines change primitives, filters on customers/products, and time granularity.

A **product-centric view** of our data is a function $\texttt{PView} : T \times P \times C \rightarrow \mathbb{N}^+ \times T$ that admits a time period T, a set of products $P \subseteq \mathcal{P}$, a set of customers $C \subseteq \mathcal{C}$, and produces a time series $D = [(a_1, t_1), \ldots, (a_n, t_n)]$ where a_i aggregates

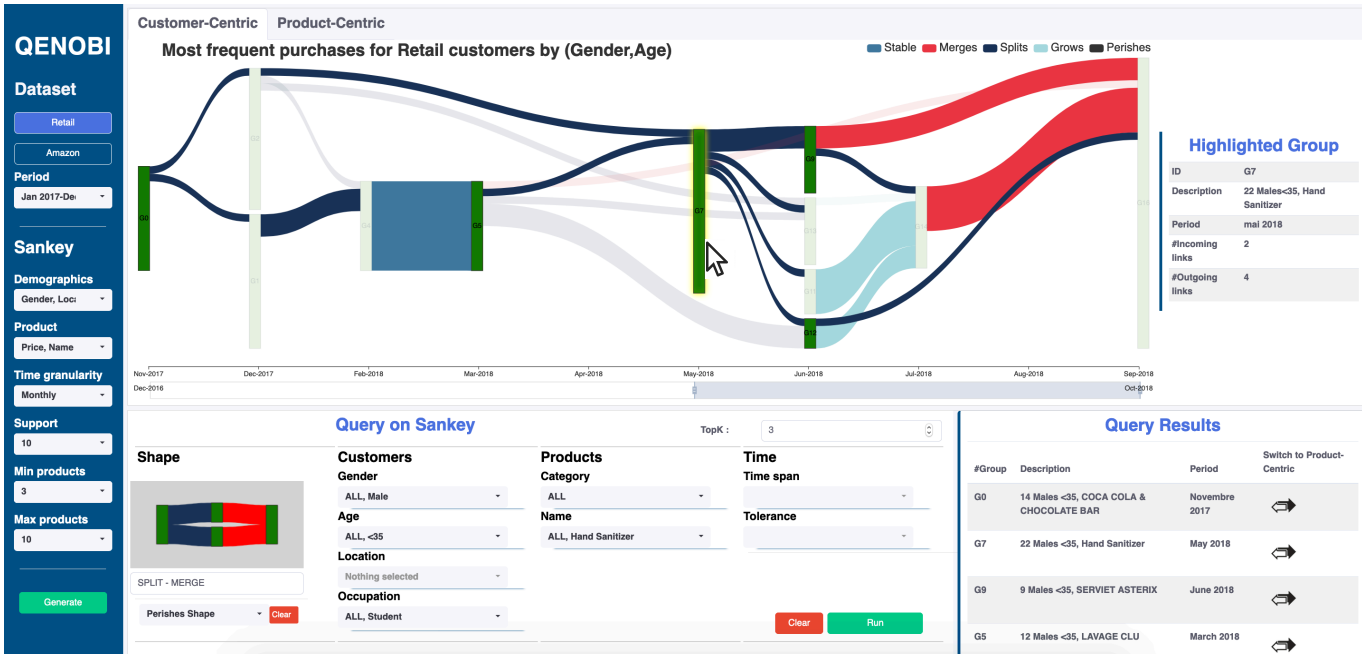

Fig. 2. Screenshot of QeNoBi showing the customer-centric view and its querying interface. A specific group can be highlighted by mouse-hovering on the Sankey diagram, e.g. the group of young adults who purchased hand sanitizers in May 2018. Groups in grey in the Sankey diagram do not satisfy the SPLIT-MERGE Shape query. The user can switch to the product-centric view from the query result panel.

Primitive	Definition	Description
$\texttt{stable}(s, S')$	$\{s' \in S' \mid (s' = s) \wedge (s'.label = s.label)\}$	a set s remains stable between periods t and t'
$\texttt{grows}(s, S')$	$\{s' \in S' \mid s \subset s' \wedge (s' \setminus s) \notin S\}$	a set s merges into a bigger set s' in period t' and whose additional members did not exist in the previous period t
$\texttt{merges}(s, S')$	$\{s' \in S' \mid s \subset s'\}$	a set s merges into a bigger set s' between periods t and t'
$\texttt{splits}(s, S')$	$\{s' \in S' \mid s' \cap s \neq \phi \wedge s' \notin \texttt{merges}(s, S')\}$	a set s splits into one or more subsets between periods t and t'
$\texttt{perishes}(s, S')$	$\nexists s' \in S' \mid s' \cap s \neq \phi$	a set s disappears in time period t'

TABLE I
CHANGE ALGEBRA PRIMITIVES FOR LABELING AND QUERYING THE GRAPH IN THE CUSTOMER-CENTRIC VIEW

the number of unique customers who purchased products in P during time period t_i. Figure 3 shows an example of the generated time series representing aggregated customer purchases for hand sanitizers that span a period of 12 months.

B. Querying Behaviors

Users can query customer behavior by means of a unified HCR (*sHape, sCope, and gRanularity*) query, noted $Q(H, C, R)$.

Querying product-centric behaviors. The result of an HCR query on a time series $D = [(a_1, t_1), \ldots, (a_n, t_n)]$ is defined as: $D^* = \arg\max_{d \in D} F(Q, d)$ where $F(Q, d)$ is a scoring function between segments $d \in D$ and H that considers similarities in terms of shape, scope, and granularity. Our Fagin-style top-k algorithm returns D^*, a set of k highest scoring segments in D with respect to the input query Q.

For instance, the query could ask to return the best matching segments that show a sharp increase in hand sanitizer purchases among the older population during a granularity of a month.

Querying customer-centric behaviors. A query Q on the behavioral graph G is composed of H, a logical expression that combines our change primitives, C and R that are defined similarly to the above. Query results are defined by: $G^* = \{g \in G \mid F(Q, g)\ \textit{is true}\}$ where g is a subgraph in G and $F(Q, g)$ is a Boolean function that returns true iff $H(g)$ is true and g satisfies C and R. Unlike for time series, Sankey diagrams are matched in a Boolean fashion. For instance, H could be written as $(\texttt{merges}(s, S_i) \rightarrow \texttt{splits}(s', S_j)) \neq \phi$ to mean that we are looking for a shape where a group of customers (e.g., older customers in a given region) merges into another group (e.g., younger customers), exhibiting similar purchase behavior, followed by a split into two groups (e.g., male customers and female customers) due to a change of interest in the products.

Our data model of our demo relies on the basket-market model to exhibit the interactions between customers and products. However our system can generalize to other data models as long as they offer two subjects with different interactions changing over time (e.g., clients/reviews, patients/treatments, agents/systems).

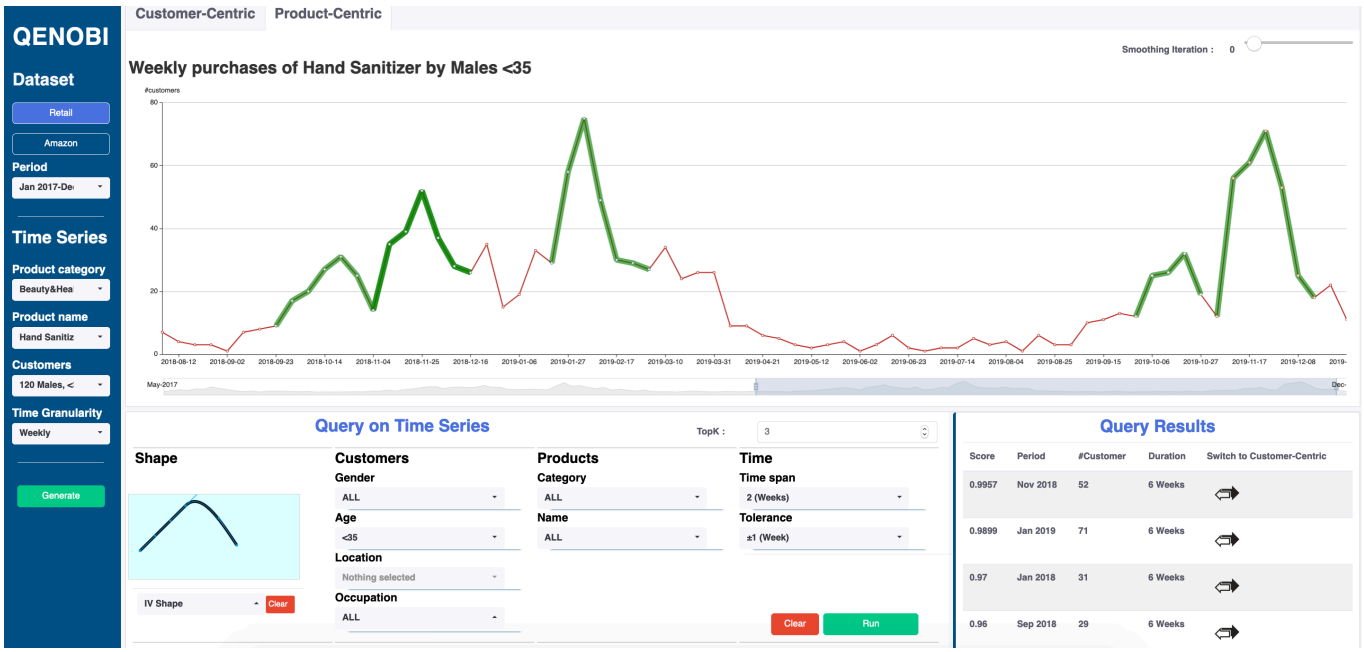

Fig. 3. Screenshot of QeNoBi showing the product-centric view and its querying interface to select customers with particular demographics, discover, and highlight corresponding top-k segments of the time series (k=3) with a specified shape (Inverted V).

IV. Demonstration Scenarios

Attendees will be invited to play the role of the marketing analyst interested in following: (1) a set of customers purchasing products with a particular behavior(Scenario 1) and (2) a set of products over a chosen period (Scenario 2). We will use two datasets: one containing customers reviewing Electronics products in Amazon, and the other from an industrial partner that represents a set of customers buying products in over 1,800 physical stores. As a first example, the attendee would choose a data set to mine groups of reviewers who rated highly products in the Electronics category. The attendee will start with a **customer-centric view** to generate the corresponding Sankey diagram as shown in Figure 2 and interact with it by specifying HCR queries. The shape could either be composed in the drawing panel or chosen from a workload. The goal is to find various types of customers: a set of customers whom behavior remained stable over time (providing a sustained view over time), or a set of customers who reviewed several different products and merged at some point in time (providing a diverse view). The attendee can choose one set of customers and switch for the corresponding product-centric view (shown in Figure 3) that provides information on how the purchases of those products evolved during the same period of time. The resulting view can be queried with a predefined set of shapes that reflect different trends: sharp increase in interest, stable interest, or sharp decrease in interest.

In the second example shown in Figure 3, the attendee will specify a set of products of interest, a set of customers and a time period, and generate a **product-centric view**. The attendee can query this view to focus on periods exhibiting sharp increase/decrease of interest in the products. The attendee can then discover the customer groups that support the top-3 segments (highlighted in green), choose one segment, and switch to a customer-centric view for that segment. The corresponding Sankey diagram will be updated automatically to match and reflect the corresponding customer-centric view.

References

[1] A. Aris, A. Khella, P. Buono, B. Shneiderman, and C. Plaisant. Timesearcher 2. *HCI Lab, University of Maryland*, 2005.

[2] G. Y. Chan, F. Chirigati, H. Doraiswamy, C. T. Silva, and J. Freire. Querying and exploring polygamous relationships in urban spatio-temporal data sets. In *ACM SIGMOD*, pages 1643–1646, 2017.

[3] P. Eichmann, F. Solleza, N. Tatbul, and S. Zdonik. Visual exploration of time series anomalies with metro-viz. In *ACM SIGMOD*, pages 1901–1904, 2019.

[4] D. Kifer, S. Ben-David, and J. Gehrke. Detecting change in data streams. In *VLDB*, pages 180–191, 2004.

[5] W. Li, X. Jin, and X. Ye. Detecting change in data stream: Using sampling technique. In *ICNC 2007*, pages 130–134, 2007.

[6] L. Luo, B. Li, I. Koprinska, S. Berkovsky, and F. Chen. Discovering temporal purchase patterns with different responses to promotions. In *CIKM'16*, page 2197–2202, 2016.

[7] L. Luo, B. Li, I. Koprinska, S. Berkovsky, and F. Chen. Tracking the evolution of customer purchase behavior segmentation via a fragmentation-coagulation process. In *IJCAI-17*, pages 2414–2420, 2017.

[8] M. Mannino and A. Abouzied. Qetch: Time series querying with expressive sketches. In *ACM SIGMOD 2018*, pages 1741–1744, 2018.

[9] R. Miglausch John. Thoughts on rfm scoring. *J. of Database Marketing*, 8(1):7, 2000.

[10] R. Neamtu, R. Ahsan, C. Lovering, C. Nguyen, E. A. Rundensteiner, and G. N. Sárközy. Interactive time series analytics powered by ONEX. In *ACM SIGMOD*, pages 1595–1598, 2017.

[11] B. Omidvar-Tehrani and S. Amer-Yahia. User group analytics survey and research opportunities. *IEEE TKDE*, 2019.

[12] T. Siddiqui, P. Luh, Z. Wang, K. Karahalios, and A. G. Parameswaran. Shapesearch: A flexible and efficient system for shape-based exploration of trendlines. In *SIGMOD 2020*, pages 51–65, 2020.

[13] M. Steiger, J. Bernard, S. Mittelstädt, H. Lücke-Tieke, D. A. Keim, T. May, and J. Kohlhammer. Visual analysis of time-series similarities for anomaly detection in sensor networks. *Comput. Graph. Forum*, 33(3):401–410, 2014.

[14] T. Uno, M. Kiyomi, H. Arimura, et al. LCM ver. 2: Efficient mining algorithms for frequent/closed/maximal itemsets. In *FIMI, Vol. 126*, 2004.

[15] K. Zoumpatianos, S. Idreos, and T. Palpanas. RINSE: interactive data series exploration with ADS+. *VLDB Endow.*, 8(12):1912–1915, 2015.