additional constraints specified by the user, restricting the set of permitted entities in the solution.

The final output of InstantEspresso is an easy-to-comprehend visualization of the relationship summarization. More precisely, as shown in the schematic overview in Figure 1 as well as in the screenshot taken from the actual system (Figure 3), the computed relatedness cores are shown together with their connections to the individual query entities, providing a concise yet comprehensive summarization of the relationship. The cognitive load on the user is controlled by satisfying a size constraint on the number of displayed cores as well as the sizes of the individual relatedness core subgraphs.

3. USE CASES

3.1 Graphical Relationship Summarization

InstantEspresso provides a graph-based summarization of the relationship between two sets of entities, permitting non-expert users to leverage the wealth of information encoded in the underlying knowledge graph. A particularly useful application of InstantEspresso lies in the summarization of relationships between real-world entities such as countries, politicians, organizations, or athletes. During preparation of a novel article, journalists can easily research the background and interaction history of the entities of interest. As an example, for an article involving the annexation of Crimea by the Russian Federation, it is useful for the journalist to research the development of the relationship between the involving countries over recent years, i. e. – depending on the focus – between the Russian Federation and the Ukraine, or, for a wider scope between the Russian Federation and neighboring states or even between the Russian Federation and Western countries. Results returned for a query on the former scenario posed to the Espresso system include event complexes such as the *Orange Revolution* and the *Russo-Georgian War*. Through the integration of rich, structured knowledge bases (YAGO and Freebase), additional constraints can be added to the user query for a more focused retrieval. Examples include the restriction of event types (e. g. to *military conflicts*, *sports events*), geographic location, or date. In addition, users can provide keywords to constrain the possible results to entities whose description (Wikipedia article) contains the specified terms.
In the final output, edges are visualized between query entities and relatedness cores (event complexes) in a form that captures the degree of involvement. This way, it is easy to see which subsets of the query entity sets are involved in which event.

3.2 Retrieval of Relevant Documents

The purely graph-based visualization of the interaction histories provides a high-level relationship summarization. In many cases, users may require further information about the displayed event complexes. To this end, InstantEspresso provides an alignment of the displayed relatedness cores with large text corpora, allowing to directly retrieve documents (web pages and news articles) involving the relevant subsets of the query entities and the computed relatedness cores. These documents provide a textual description of the event complexes. Document retrieval is achieved by the integration of two external data sources, the ClueWeb09 corpus with annotated Freebase entities [5], as well as STICS [6], an entity-search engine over a continuously updated stream of news articles. The latter provides up-to-date news articles (currently over 1.5 million) from more than 100 feeds. As soon as a user has specified the query entities of interest, InstantEspresso first computes the graph-based relationship summarization, and subsequently queries the two corpora for documents containing entities from both sets as well as from the event complexes. For the example depicted in Figure 1, results obtained from the STICS corpus correspond to news articles involving at least one politician from each query set. The documents are ranked by a combination of traditional measures such as PageRank as well as the number of annotated relevant entities.

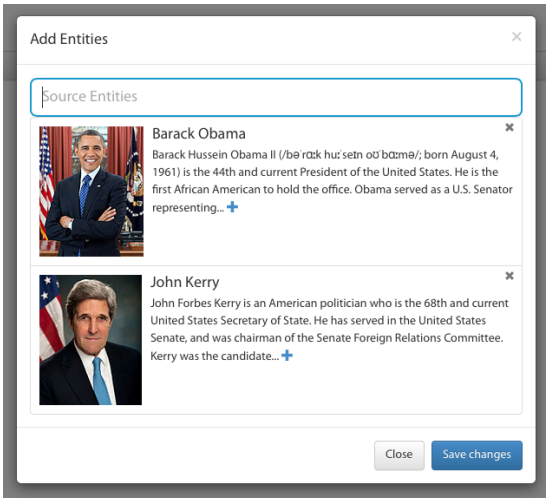

Figure 2: User Interface for Query Specification

3.3 Aggregated Relationship Summarization

In addition to providing summarizations in the form of the most important event complexes in the form of informative subgraphs, InstantEspresso also provides the user with higher-order overviews, offering an aggregated view of the interaction history between the query entities.

Relationship Heatmap. Especially for large query sets – e. g. *US politicians*, *Oil Companies*, etc. – it makes sense to first identify subsets of interrelated entities from both query sets and subsequently display the relationship summarizations for the respective subsets selected by the user. A visually appealing way to identify these subsets are heatmaps indicating the strength of interactions. Here, entities are first clustered based on their co-occurrence and then arranged as the columns and rows of a matrix representation. The matrix cells are colored based on the number of co-occurrences of the entities corresponding to row and column. Selection of a cell (submatrix) then leads to focus on the respective entity subsets for the subsequent relationship summarization by the graphical display of relatedness cores.

Topical Classification. As an additional feature, InstantEspresso aggregates the computed event complexes into a coarse-grained overview in the form of topics (e. g. sports, politics, economics, etc.). For this purpose, we have trained a multinomial classifier, assigning each article to one or more topics from a predefined set. After computing the most characteristic event complexes, InstantEspresso displays an aggregated view of the relationship based on the topics associated with the central events.

Temporal Relationship Analysis. Another insightful aggregated view is achieved by the temporal alignment of event complexes. For this purpose, InstantEspresso assigns the extracted relatedness cores to the (discretized, e. g. yearly) time axis, based on the event timespan as recorded in the Freebase knowledge base. Each time interval receives a score, based on the number of event complexes

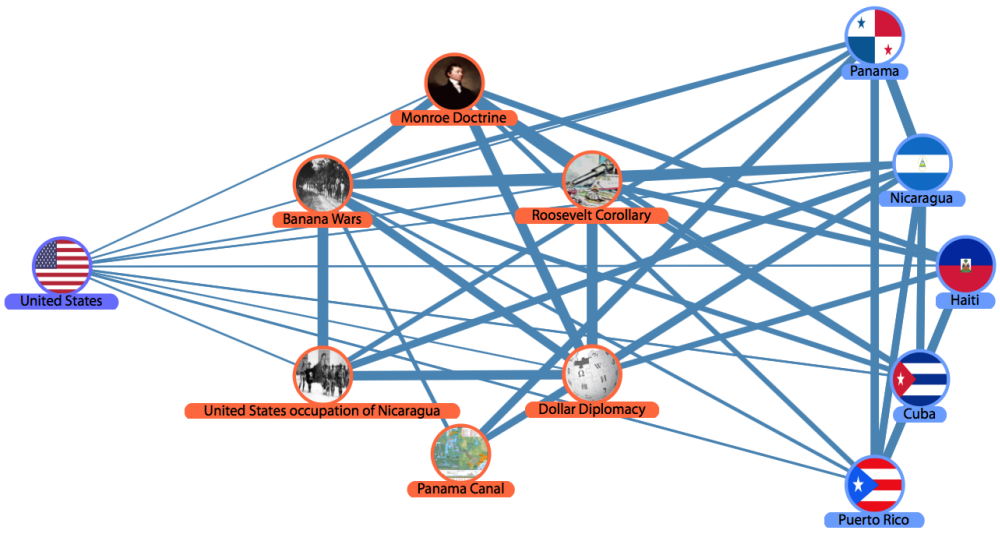

Figure 3: Relatedness Core between the United States and countries from the Americas

relevant to the interval, weighted by the degree of involvement of the individual query entities and the overall event importance (derived from structural measures such as degree centrality or PageRank, as well as popularity measured by number of pageviews). The resulting visualization gives an overview over the periods of time of highest importance to the queried relationship.

4. SYSTEM DEMONSTRATION

Hardware. InstantEspresso over the Espresso Knowledge Graph has modest hardware requirements. 8 GB of main memory are sufficient to represent the structural information of the graph used for computing the random walk scores for candidate event ranking. InstantEspresso will be demonstrated on an out-of-the-box Apple MacBook Pro, equipped with an Intel Core i7@2.3Ghz Quad Core CPU and 16 gigs of main memory.

Query Processing. The time required to answer a query is independent of the number of query entities specified by the user, since we use a random walk approach by power iteration with a fixed number of iterations. Thus, the amount of computation is constant, only the starting probabilities differ across queries. InstantEspresso provides answers to each query after roughly 20 seconds.

User Interaction. Users can interact with InstantEspresso by specifying two sets of query entities and inspecting the computed results. Entities can be specified either enumeratively or by selecting one or more predefined entity sets. The query specification interface is shown in Figure 2. In addition, users can specify a time interval of interest. InstantEspresso then first displays aggregated information, such as the heatmap and temporal overviews discussed in Section 3.3. Afterwards, the graphical representation is shown to the user (see Figure 3 for an example). Users can click on entities to see a short description of the entities. Below the graphical representation, InstantEspresso provides a list of relevant retrieved documents from ClueWeb and STICS.

5. CONCLUSION

We demonstrate InstantEspresso, a system for interactive analysis of relationships between two sets user-specified entities over a knowledge graph. Our system summarizes relationships by means of extracting important, relevant, and coherent thematic complexes corresponding to real-world events. These thematic complexes are modeled as size-constrained, dense subgraphs that are well connected with the query entities in the knowledge graph. Our system provides an interface to specify the query entities of interest, based on this selection extracts informative subgraphs from the knowledge graph, and displays the extracted results in a visually appealing graph visualization together with aggregated information and relevant news and web documents. An important target application of InstantEspresso is its use to quickly gain background information about the interaction history of entities, e. g. for a journalist preparing an article involving the query entities.

6. REFERENCES

[1] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. DBpedia: A Nucleus for a Web of Open Data. In *The Semantic Web*, LNCS volume 4825, 2007.

[2] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge. In *SIGMOD'08*.

[3] C. Faloutsos, K. S. McCurley, and A. Tomkins. Fast Discovery of Connection Subgraphs. In *KDD'04*.

[4] L. Fang, A. D. Sarma, C. Yu, and P. Bohannon. REX: Explaining Relationships between Entity Pairs. *PVLDB*, 5(3):241–252, 2011.

[5] E. Gabrilovich, M. Ringgaard, and A. Subramanya. FACC1: Freebase Annotation of ClueWeb Corpora, Version 1, 2013.

[6] J. Hoffart, D. Milchevski, and G. Weikum. STICS: Searching with Strings, Things, and Cats. In *SIGIR'14*.

[7] J. Hoffart, S. Seufert, D. B. Nguyen, M. Theobald, and G. Weikum. KORE: Keyphrase Overlap Relatedness for Entity Disambiguation. In *CIKM'12*.

[8] J. Hoffart, F. M. Suchanek, K. Berberich, and G. Weikum. YAGO2: A Spatially and Temporally Enhanced Knowledge Base from Wikipedia. *Artificial Intelligence*, 2013.

[9] G. Kasneci, S. Elbassuoni, and G. Weikum. MING: Mining Informative Entity-Relationship Subgraphs. In *CIKM'09*.

[10] D. Milne and I. H. Witten. An Effective, Low-Cost Measure of Semantic Relatedness Obtained from Wikipedia Links. In *WIKIAI'08*. AAAI, 2008.

[11] H. Tong and C. Faloutsos. Center-Piece Subgraphs: Problem Definition and Fast Solutions. In *KDD'06*.