eBay Shopbot

Conversational AI Recommendation System
eBay Shopbot Presentation Origin Story

Presentation based on “eBay ShopBot: Graph-powered Conversational Commerce”


eBay ShopBot: Graph-powered Conversational Commerce, GraphConnect (2017, Oct. 23-24) [Video].
One of the world’s largest and most vibrant marketplaces

- **$20B** GMV
- **1.1B** Live listings
- **80%** Items sold as new
- **67%** Transactions that ship for free (US, UK, DE)
- **12M** New listings added via mobile per week
- **60%** Platform GMV touched by mobile

**Takeaway: Scale, Volume, Variety, and Velocity Matters**

Data as of Q1 2017
## 2017 Fun Facts: Velocity

### Stats by Region

<table>
<thead>
<tr>
<th>Region</th>
<th>Frequency of product purchases via desktop and mobile</th>
<th>Takeaway: Behavior is complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>A tool is purchased every 11 sec&lt;br&gt;A smartphone is purchased every 5 sec&lt;br&gt;A watch is purchased every 4 sec</td>
<td>A smartphone is purchased every 5 sec&lt;br&gt;A watch is purchased every 4 sec&lt;br&gt;A makeup product is purchased every 3 sec&lt;br&gt;A car part is purchased every 1 sec&lt;br&gt;An appliance is purchased every 8 sec&lt;br&gt;A tire is purchased every 16 sec&lt;br&gt;A tablet is purchased every 3 sec&lt;br&gt;A Lego is purchased every 18 sec</td>
</tr>
<tr>
<td>UK</td>
<td>A makeup product is purchased every 3 sec&lt;br&gt;A car part is purchased every 1 sec&lt;br&gt;An appliance is purchased every 8 sec</td>
<td>A makeup product is purchased every 3 sec&lt;br&gt;A car part is purchased every 1 sec&lt;br&gt;An appliance is purchased every 8 sec</td>
</tr>
<tr>
<td>DE</td>
<td>A tire is purchased every 16 sec&lt;br&gt;A tablet is purchased every 3 sec&lt;br&gt;A Lego is purchased every 18 sec</td>
<td>A wedding item is purchased every 26 sec&lt;br&gt;A home decor item is purchased every 14 sec&lt;br&gt;A car or truck part is purchased every 4 sec</td>
</tr>
<tr>
<td>AU</td>
<td>A wedding item is purchased every 26 sec&lt;br&gt;A home decor item is purchased every 14 sec&lt;br&gt;A car or truck part is purchased every 4 sec</td>
<td>A wedding item is purchased every 26 sec&lt;br&gt;A home decor item is purchased every 14 sec&lt;br&gt;A car or truck part is purchased every 4 sec</td>
</tr>
</tbody>
</table>

---

**Takeaway:** Behavior is complex.
eBay Shopbot - Natural Language Understanding

NLU is not a rule based search engine – Personalization is KEY!

Personal Shopping Assistant

Powered By AI

Developed at eBay’s NPD Group

Conversational Commerce bridges the gap between stateless search engine and a shopper’s actual intent.

eBay’s goal is to be as close as possible using AI tech such as Natural Language Understanding, Knowledge Graphs and Computer Vision.

Shopbot launched in 2016 on facebook messenger platform. Now also available thru Google Home.

Developed at eBay’s NPD Group

Shopbot launched in 2016 on facebook messenger platform. Now also available thru Google Home.

Conversational Commerce bridges the gap between stateless search engine and a shopper’s actual intent.

eBay’s goal is to be as close as possible using AI tech such as Natural Language Understanding, Knowledge Graphs and Computer Vision.

Ask eBay to find me a Kate Spade bag

Sure. Getting eBay.

Great. Do you want a color that's black, brown, or something else?

Got it. Looking for a leather, canvas, or something else?

Got it. Looking for a shoulder bag, satchel, or something else?

I looked through 35 options on eBay and the best deal is $294.78. I have more info on this deal. Is it okay if I send that to your phone?

Developed at eBay’s NPD Group

Shopbot launched in 2016 on facebook messenger platform. Now also available thru Google Home.

Conversational Commerce bridges the gap between stateless search engine and a shopper’s actual intent.

eBay’s goal is to be as close as possible using AI tech such as Natural Language Understanding, Knowledge Graphs and Computer Vision.

Ask eBay to find me a Kate Spade bag

Sure. Getting eBay.

Great. Do you want a color that's black, brown, or something else?

Got it. Looking for a leather, canvas, or something else?

Got it. Looking for a shoulder bag, satchel, or something else?

I looked through 35 options on eBay and the best deal is $294.78. I have more info on this deal. Is it okay if I send that to your phone?
Why – did eBay build their own Shopbot?

- 3rd Party Bot Frameworks
  - Tuned for general purpose bots
    - Intent detection (e.g. weather, flight schedules, …)
    - Entity extraction (e.g. number, temperature, …)
  - Limited non-linear conversation support
  - Coarse grained bot memory
  - No inherent gain with experience
  - Existing offerings in various states of maturity
    - Scalability (data size, complexity, and speed)
    - Capabilities, API’s, tools, extensibility…

What – is eBay trying to solve with Shopbot?

- Conversational Commerce not search
- Dialog systems
- What’s the next best question to ask
- How do we build the learning system that humans gain inherently with experience
- Can collaborative inferencing be used

Why - is it hard to build a Shopbot

Most production chatbots have curated dialog flows

- It’s easy for restricted domains like flight booking bots
- With 20k+ categories of products with 150K+ attributes, it’s hard to get right and scale!
- Need a solution which scale for dialog conversations on thousands of products
- Different interaction / input formats such as voice and image
- Speed and effectiveness at (any) scale is required by users

Why - is Shopbot important to eBay

- Unreasonable Effectiveness of Data — Halevy, Norvig, Pereria
- eBay has seen “almost” all queries that are important
- No production system runs dialogs just based on Deep Learning models

In other terms... To sell more stuff!
My husband needs some new black leather dress shoes but I want to spend less than $80, what do you have?
How eBay addressed the need for continuous NLU Shopbot functionality

- Build a probabilistic inference graph
  - Data comes from eBay.com user behavioral patterns and other sources
  - When a user is looking for a “running shoes” what’s the most important aspect of the shoe that they care about (maybe NLU style query flow here with next questions)

- Enter GraphDB (Neo4j) for eBay Shopbot knowledge cache
  - Tried RDBMS
  - Needed query / speed / schema flexibility / production system capabilities provided by the Neo4j property graph database technology and tools
eBay ShopKnowledge Workflow at a Glance

Data Sources

Query Understanding

Trends

Price Prediction

Aspect Reco

Entity Extraction

Knowledge Graph

Microservices

Apache Airflow

Spark

neo4j

Google Cloud Platform

Amazon Web Services

LinkedIn

Intuit

Wikipedia

Entity Extraction

Product Knowledge
Data Sources

- Latest wikidata stats
- Items / txn items
- Buyer behavioral data...
- WiW – What’s It Worth
- And...

Graph DB Characteristics and High Level Functionality

- > 0.5B nodes
- > 16B relationships
- Functionality
  - Probabilistic graphical models
  - Supervised and semi-supervised models
Under the Shopbot hood ...

- Probabilistic inferencing on behavioral patterns from past users
- Almost like Expert system navigation
  - Parse, categorize, group, contextualize natural language
- Easy to combine World Knowledge to augment user behavior data with external datasets:
  - Curation to augment data
  - Machine learning models to augment data
- ML model as a cache in Neo4j
Ask for shoes by color – likely to be a woman or man?

Basis to “ask the next best question”

e.g. F: *Brand* question / suggestion

M: *Color* question / suggestion
An expanded look at part of a Knowledge Graph

Knowledge Graph Encapsulates Shopping Behavior
An expanded look at part of a Knowledge Graph

Looked simple. Why need queries and algorithms?
An expanded look at part of a Knowledge Graph

*What's the “next best question” (a.k.a. query the graph)*

5th Annual Big Data & Business Analytics Symposium - March 22-23, 2018
Transfer Learning From Interaction

Example Product Conversation

I am looking for the eggplant foamposites.

- Color: Purple
- Material: Foamposite
- Category: Athletic Shoes
- Product: Eggplant Foamposite Sneakers
- Brand: Nike
- Style: Basketball Shoes
- Release Date: 2009

Which one would you pick?

"Knowledge" added to graph

I am looking for the **eggplant iphone case**.

Data mine from the graph that eggplant == purple
Derived meaning from data (unsupervised)
Why Neo4j?!

- Graph database is the right choice to store Knowledge Graphs
- Neo4j is battle tested and fast!
- New data sources are easy to integrate to extend Graph inferencing ability
- Great set of support tools, to name a few –
  - Interactive browser
  - In-Database Procedures
  - Bulk imports
  - Graph-algorithms

*Walter 1st tried to build with RDBMS*
Graphs can do some amazing things

Walter: “Say data joins one more time”

eBay “initial tried RDBMS for knowledge graph”. What happened?

**Find all direct reports and how many people they manage, up to 3 levels down**

**SQL Query**

```
SELECT t.directReports AS directReports, count(*) AS count
FROM (...
```

**Graph DB Query**

(uses Cypher Query Language)

```
MATCH (sub)-[r:REPORTS_TO*0..3]->(boss),
    (report)-[r:REPORTS_TO*1..3]->(sub)
WHERE boss.name = "John Doe"
RETURN sub.name AS Subordinate,
    count(report) AS Total
```

Graph database queries can do amazing things fast, but graph algorithms add a whole new level of capability…
... and recommendation engines, and ...

- Business logic and interference, and “creative juices” in graph.
- Interactive search
- Search science pieces as graph
- Everything except backend source systems (origin indexes, etc.)
- Semi-supervised techniques like label propagation
# Neo4j - Some Graph Algorithms

## Path Finding and Traversal Algorithms

<table>
<thead>
<tr>
<th>Algorithm Type</th>
<th>What It Does</th>
<th>Example Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parallel Breadth-First Search (BFS)</strong></td>
<td>Traverses a tree data structure by fanning out to explore the nearest neighbors and then their sub-level neighbors. It’s used to locate connections and is a precursor to many other algorithms. BFS is preferred when the tree is less balanced or the target is closer to the starting point. It can also be used to find the shortest path between nodes or avoid recursive processes of DFS.</td>
<td>BFS can be used to locate neighbor nodes in peer-to-peer networks like BitTorrent, GPS systems to pinpoint nearby locations and social network services to find people within a specific distance.</td>
</tr>
<tr>
<td><strong>Parallel Depth-First Search (DFS)</strong></td>
<td>Traverses a tree data structure by exploring as far as possible down each branch before backtracking. It’s used on deeply hierarchical data and is a precursor to many other algorithms. DFS is preferred when the tree is more balanced or the target is closer to an endpoint.</td>
<td>DFS is often used in gaming simulations where each choice or action leads to another, expanding into a tree-shaped graph of possibilities. It will traverse the choice tree until it discovers an optimal solution path (e.g., win).</td>
</tr>
<tr>
<td><strong>Single-Source Shortest Path</strong></td>
<td>Calculates a path between a node and all other nodes whose summed value (weight of relationships such as cost, distance, time or capacity) to all other nodes are minimal.</td>
<td>Single-Source Shortest Path is often applied to automatically obtain directions between physical locations, such as driving directions via Google Maps. It’s also essential in logical routing such as telephone call routing (least cost routing).</td>
</tr>
<tr>
<td><strong>All-Pairs Shortest Path</strong></td>
<td>Calculates a shortest path forest (group) containing all shortest paths between the nodes in the graph. Commonly used for understanding alternate routing when the shortest route is blocked or becomes suboptimal.</td>
<td>All-Pairs Shortest Path can be used to evaluate alternate routes for situations such as a freeway backup or network capacity. It’s also key in logical routing to offer multiple paths, for example, call routing alternatives.</td>
</tr>
<tr>
<td><strong>Minimum Weight Spanning Tree (MWST)</strong></td>
<td>Calculates the paths along a connected tree structure with the smallest value (weight of the relationship such as cost, time or capacity) associated with visiting all nodes in the tree. It’s also employed to approximate some NP-hard problems such as the traveling salesman problem and randomized or iterative rounding.</td>
<td>MWST is widely used for network designs: least cost logical or physical routing such as laying cable, fastest garbage collection routes, capacity for water systems, efficient circuit designs and much more. It also has real-time applications with rolling optimizations such as processes in a chemical refinery or driving route corrections.</td>
</tr>
</tbody>
</table>
## Centrality Algorithms

<table>
<thead>
<tr>
<th>Algorithm Type</th>
<th>What It Does</th>
<th>Example Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PageRank</strong></td>
<td>Estimates a current node’s importance from its linked neighbors and then again from their neighbors. A node’s rank is derived from the number and quality of its transitive links to estimate influence. Although popularized by Google, it’s widely recognized as a way of detecting influential nodes in any network.</td>
<td>PageRank is used in quite a few ways to estimate importance and influence. It’s used to suggest Twitter accounts to follow and for general sentiment analysis. PageRank is also used in machine learning to identify the most influential features for extraction. In biology, it’s been used to identify which species extinctions within a food web would lead to biggest chain-reaction of species death.</td>
</tr>
<tr>
<td><strong>Degree Centrality</strong></td>
<td>Measures the number of relationships a node (or an entire graph) has. It’s broken into indegree (flowing in) and outdegree (flowing out) where relationships are directed.</td>
<td>Degree Centrality looks at immediate connectedness for uses such as evaluating the near-term risk of a person catching a virus or hearing information. In social studies, indegree of friendship can be used to estimate popularity and outdegree as gregariousness.</td>
</tr>
<tr>
<td><strong>Closeness Centrality</strong></td>
<td>Measures how central a node is to all its neighbors within its cluster. Nodes with the shortest paths to all other nodes are assumed to be able to reach the entire group the fastest.</td>
<td>Closeness centrality is applicable in a number of resources, communication and behavioral analysis, especially when interaction speed is significant. It has been used to identifying the best location of new public services for maximum accessibility. In social analysis, it can be used to find people with the ideal social network location for faster dissemination of information.</td>
</tr>
<tr>
<td><strong>Betweenness Centrality</strong></td>
<td>Measures the number of shortest paths (first found with BFS) that pass through a node. Nodes that most frequently lie on shortest paths have higher betweenness centrality scores and are the bridges between different clusters. It is often associated with the control over the flow of resources and information.</td>
<td>Betweenness Centrality applies to a wide range of problems in network science and can be used to pinpoint bottlenecks or likely attack targets in communication and transportation networks. In genomics, it has been used to understand the control certain genes have in protein networks for improvements such as better drug-disease targeting. Betweenness Centrality has also be used to evaluate information flows between multiplayer online gamers and expertise sharing communities of physicians.</td>
</tr>
</tbody>
</table>

Determine the importance of distinct nodes in a network of connected data.
# Community Detection Algorithms (a.k.a. clustering and partitioning)

<table>
<thead>
<tr>
<th>Algorithm Type</th>
<th>What It Does</th>
<th>Example Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Label Propagation</strong></td>
<td>Spreads labels based on neighborhood majorities as a means of inferring clusters. This extremely fast graph partitioning requires little prior information and is widely used in large-scale networks for community detection. It's a key method for understanding the organization of a graph and is often a primary step in other analysis.</td>
<td>Label Propagation has diverse applications from understanding consensus formation in social communities to identifying sets of proteins that are involved together in a process (functional modules) for biochemical networks. It's also used in semi- and unsupervised machine learning as an initial preprocessing step.</td>
</tr>
<tr>
<td><strong>Strongly Connected</strong></td>
<td>Locates groups of nodes where each node is reachable from every other node in the same group following the direction of relationships. It's often applied from a depth-first search.</td>
<td>Strongly Connected is often used to enable running other algorithms independently on an identified cluster. As a preprocessing step for directed graphs, it can help quickly identify disconnected groups. In retail recommendations, it can help identify groups with strong affinities that then can be used for suggesting commonly preferred items to those within that group who have not yet purchased the item.</td>
</tr>
<tr>
<td><strong>Union-Find / Connected Components / Weakly Connected</strong></td>
<td>Finds groups of nodes where each node is reachable from any other node in the same group, regardless of the direction of relationships. It provides near constant-time (independent of input size) operations to add new groups, merge existing groups and determine whether two nodes are in the same group.</td>
<td>Union-Find / Connected Components is often used in conjunction with other algorithms, especially for high-performance grouping. As a preposing step for undirected graphs, it can help quickly identify disconnected groups.</td>
</tr>
<tr>
<td><strong>Louvain Modularity</strong></td>
<td>Measures the quality (i.e., presumed accuracy) of a community grouping by comparing its relationship density to a suitably defined random network. It's often used to evaluate the organization of complex networks, in particular, community hierarchies. It's also useful for initial data preprocessing in unsupervised machine learning.</td>
<td>Louvain is used to evaluate social structures in Twitter, LinkedIn and YouTube. It's used in fraud analytics to evaluate whether a group has just a few bad behaviors or is acting as a fraud ring that would be indicated by a higher relationship density than average. Louvain revealed a six-level customer hierarchy in a Belgian telecom network.</td>
</tr>
</tbody>
</table>
# Neo4j - Some Graph Algorithms

**Community Detection Algorithms (a.k.a. clustering and partitioning)**

| Algorithm Type                  | What It Does                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Example Uses                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|---------------------------------|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Local Cluster Coefficient is important to estimating resilience by understanding the likelihood of group coherence or fragmentation.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| **Local Clustering Coefficient / Node Clustering Coefficient**                                                                          | For a particular node, it quantifies how close its neighbors are to being a clique (every node is directly connected to every other node). For example, if all your friends knew each other directly, your local clustering coefficient would be 1. Small values for a cluster would indicate that although a grouping exists, the nodes are not tightly connected.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Analysis of a European power grid using this method found that clusters with sparsely connected nodes were more resilient against widespread failures.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| **Triangle-Count and Average Clustering Coefficient**                                                                                  | Measures how many nodes have triangles and the degree to which nodes tend to cluster together. The average clustering coefficient is 1 when there is a clique, and 0 when there are no connections. For the clustering coefficient to be meaningful it should be significantly higher than a version of the network where all of the relationships have been shuffled randomly.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | The Average Clustering Coefficient is often used to estimate whether a network might exhibit "small-world" behaviors which are based on tightly knit clusters. It's also a factor for cluster stability and resiliency. Epidemiologists have used the average clustering coefficient to help predict various infection rates for different communities.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
Some related and interesting material

**eBay Presentation / Article** (one of many)
Ajinkya Kale and Anju Vasta, New Product Development, eBay

- https://www.youtube.com/watch?v=hRpmLeJjx-Y
- https://www.youtube.com/playlist?list=PL9HI4pk2FsvVdYIoyksOyAMyDEDYv6O4K

**Why Knowledge Graphs Are Foundational to Artificial Intelligence**
Jim Webber, Chief Data Scientist, Neo4j, March 20, 2018

https://www.datanami.com/2018/03/20/why-knowledge-graphs-are-foundational-to-artificial-intelligence

**Graph Analytics: Graph Algorithms inside Neo4j**
Amy Hodler and Michael Hunger, Neo4j, January 26, 2018

https://www.youtube.com/watch?v=y10Bt7OkCRM
Thank you! Neo4j

Remember – “The world is connected. Value comes from being able to make and understand the connections”*

* A wise quote from my wonderful wife spouse before saying: “I know you did it, and don’t do it again.”