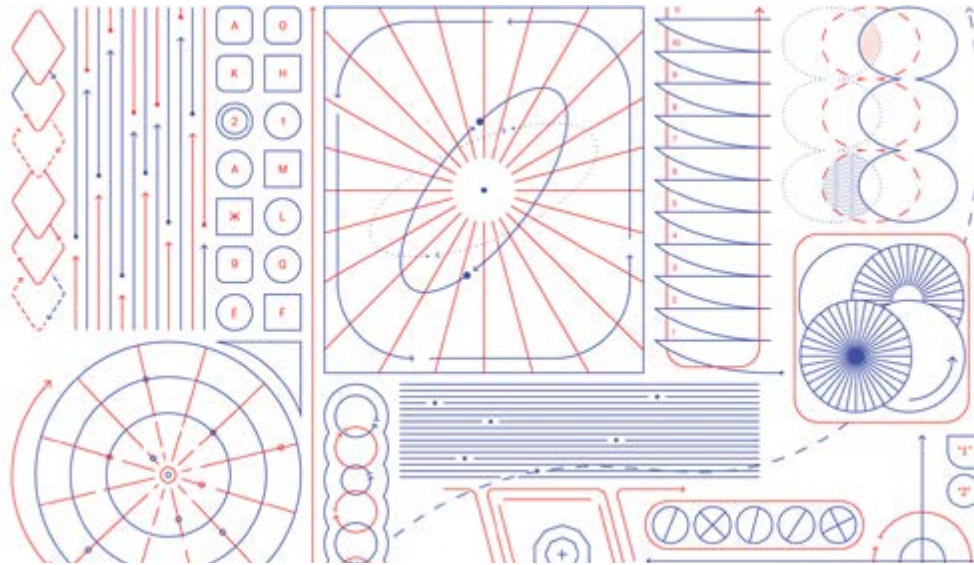


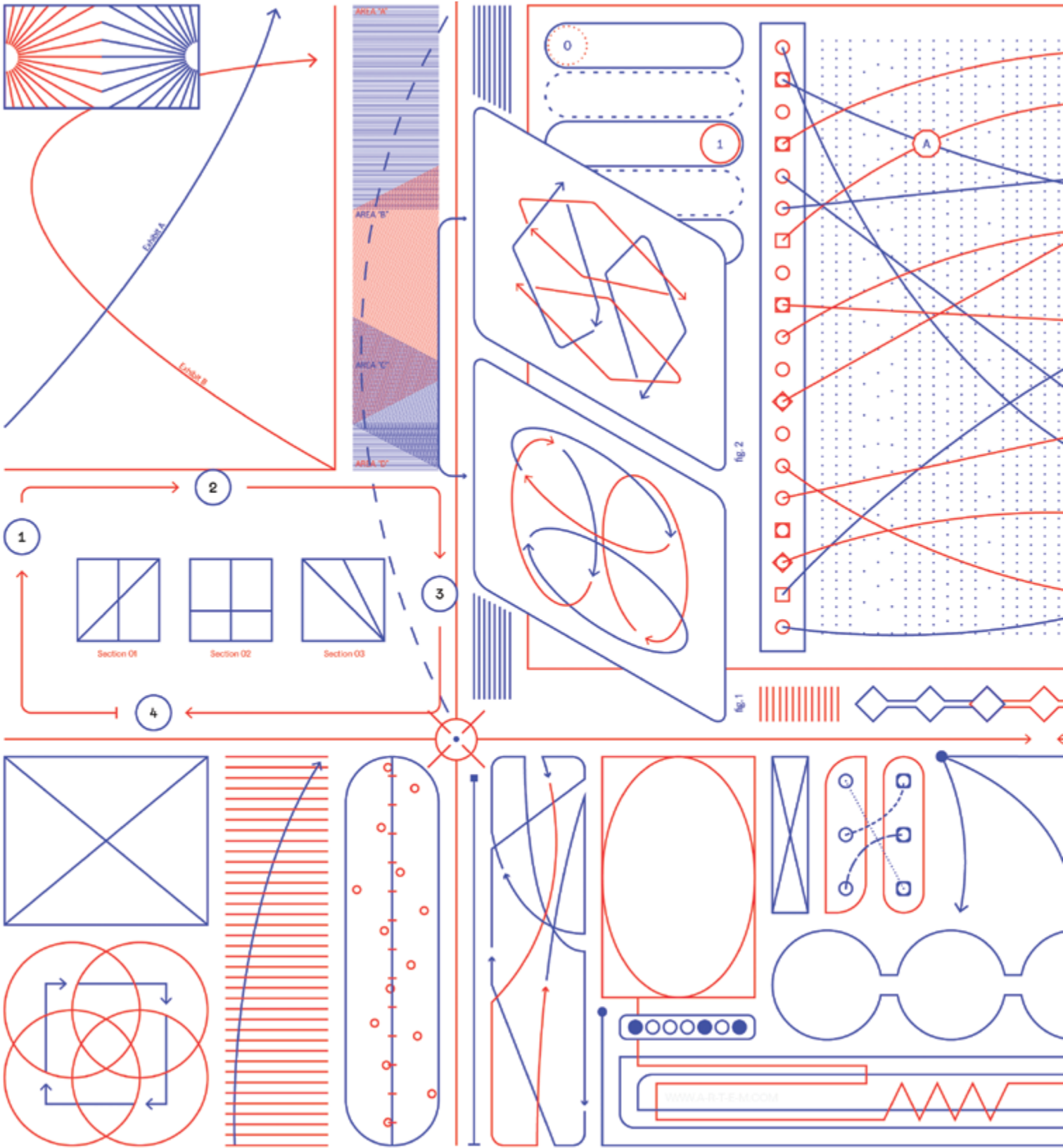


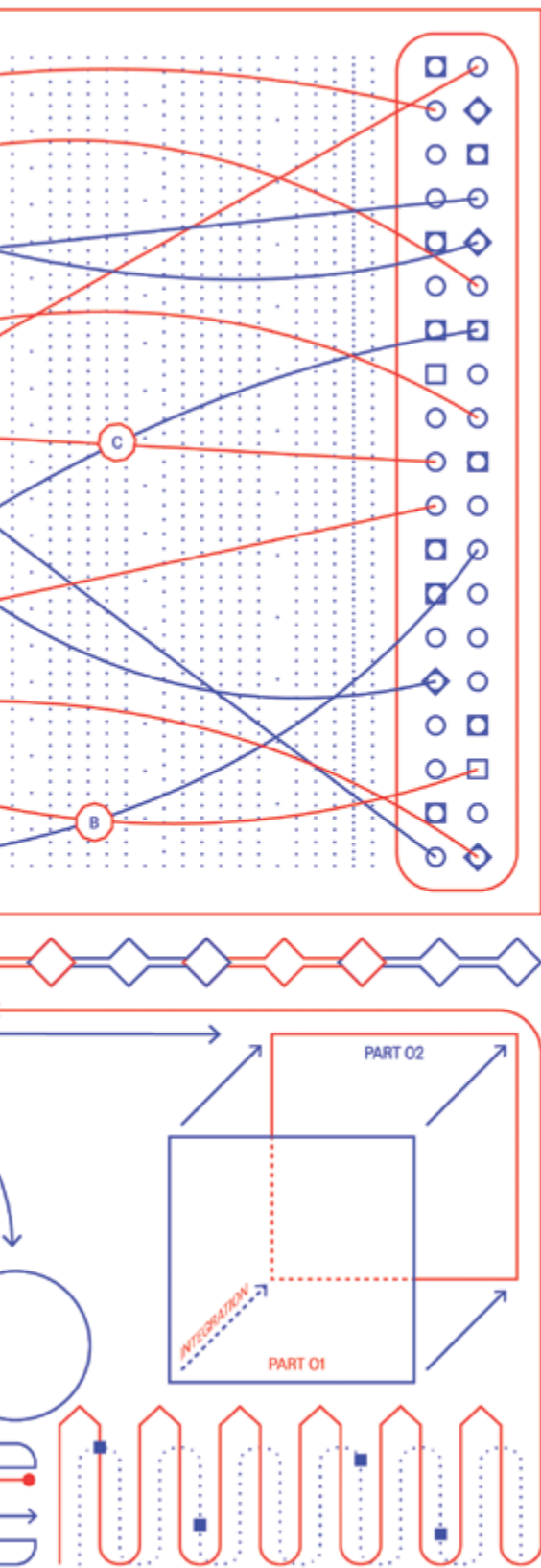
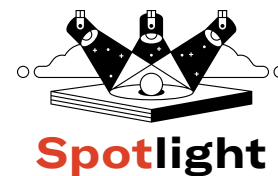
Digital Literacy



The Next Great Digital Advantage

by Vijay Govindarajan and N. Venkat Venkatraman





The Next Great Digital Advantage

Smart businesses are using datagraphs to reveal unique solutions to customer problems.



AUTHORS

Vijay Govindarajan
Professor, Dartmouth's
Tuck School of Business

N. Venkat Venkatraman
Professor, Boston University's
Questrom School of Business



OF THE 4,000 products Amazon sells every minute, approximately 50% are presented to customers by its personalized recommendation engine. When you visit the site, its algorithms select an assortment of products from about 353 million items and arrange them for you according to what they predict you will want at that precise moment. These recommendations are powered by Amazon's ever-evolving purchase graph, which is a digital representation

of real-world "entities"—anything about which it stores information, such as customers, products, purchases, events, and places—and the relationships and interrelationships among them. Amazon's purchase graph connects purchase history with browsing data on the site, viewing data on Prime Video, listening data on Amazon Music, and data from Alexa-enabled devices. Its algorithms use collaborative filtering—incorporating factors such as diversity (how dissimilar the recommended items are); serendipity (how



Datagraphs capture how people work, play, learn, socialize, transact, travel, and do any other activity that can be associated with commerce.

surprising they are); and novelty (how new they are)—to generate some of the most sophisticated recommendations on the planet. Thanks to its rich data and industry-leading personalization, Amazon now owns 40% of the U.S. e-commerce market; its closest rival, Walmart, has a market share of only 7%.

To compete with Amazon, in April 2021 Google announced its Shopping Graph, an AI-enhanced model that recommends products to users as they search. More than a billion people research products on Google each day, and Shopping Graph connects them with more than 24 billion listings from millions of merchants across the web. It builds on Google's unparalleled Knowledge Graph, which captures information about the entities in its vast network and the relationships among them, including structured and unstructured data from Android, voice and image search, Chrome browser extensions, Google Assistant, Gmail, Photos, Maps, YouTube, Google Cloud, and Google Pay. With its Shopping Graph—which lets 1.7 million merchants feature relevant listings across Google using simple but interlinked tools—Google is ready to meet Amazon's challenge.

Datagraphs like Amazon's and Google's rely on product-in-use data—that is, data on the behavior of customers as they use a platform or a product—to capture the connections, relationships, and interrelationships between a company and its customers. The datagraph concept is inspired by social network and graph theory, wherein a social graph is defined as a representation of the interconnections among individuals, depicted as nodes, and

the relationships among them—with friends, colleagues, supervisors, and so on—represented as links. The concept derives from the work of the social psychologist Stanley Milgram, and over the past two decades, it has provided a useful lens for analyzing the structure and dynamics of organizations, industries, markets, and societies. Facebook popularized the digital social graph in 2007 when it introduced Facebook Platform, a tool that allowed developers to build applications that were integrated into the site's information flow and connections of relationships.

Leading technology companies are using datagraphs to personalize customer recommendations, update products, optimize advertising, and more. The most successful examples—which include Amazon's purchase graph, Google's search graph, Facebook's social graph, Netflix's movie graph, Spotify's music graph, Airbnb's travel graph, Uber's mobility graph, and LinkedIn's professional graph—leverage the ongoing collection of customer engagement data, coupled with proprietary algorithms, to outcompete rivals in every way, from product creation to user experience.

This article discusses how companies can learn from the best practices of datagraph leaders to gain new competitive advantage.

Data Network Effects

To understand datagraphs, we first need to understand *data network effects*, which occur when data generated by users as they engage with a product or service makes it more valuable for other

users. Unlike direct network effects, in which the value of a service grows as additional users join (as with Facebook or LinkedIn), data network effects do not require increasing numbers of users to enhance the value of the network. Instead, the continued engagement of current users generates broader and deeper product-in-use data, which allows algorithms to generate ever-improving results. For example, every one of Google's 2 trillion annual searches helps the company enrich its Knowledge Graph and improve its search engine, which generates better and better search results for users. By contrast, if users stop engaging on the platform, it becomes stale and less useful.

Datagraphs are not static; they do not reflect information at a snapshot in time. They are dynamic, reflecting what data scientists refer to as *data in motion*. That's partly why it is impossible to manually draw a datagraph. Technology is needed to gather and interpret in real time the data on the millions of units of a company's products that consumers worldwide may be engaging with at any given moment.

Datagraph Success Factors

Datagraph leaders gather customer behavioral data and quickly incorporate what they learn to improve every aspect of their products and services. They constantly refine how they classify and label product data and uncover relationships among entities so that algorithms can better group offerings for personalized recommendations. And they continually update their algorithms so that the personalized recommendations are



A New View of Competitive Advantage

Datagraphs are redefining how companies win.

TRADITIONAL ADVANTAGES	ADVANTAGES FROM DATAGRAPHS
Scale advantage comes from physical assets	Scale advantage comes from product-in-use data
Scope expansion is based on product-market extensions	Scope expansion comes from datagraph and analytics competencies
Advantage comes from direct and indirect network effects	Advantage comes from data network effects along with direct and indirect network effects
Data enhances operational efficiency	Data provides real-time insights into customer behavior and needs
Data is stored as independent records in separate functional databases	Data is stored in integrated graph-based databases
Analytics are used to improve operations and marketing effectiveness	Analytics are used to build competitive differentiation
Firms design and deliver products and services	Firms solve customer problems with unique solutions derived from digital learning

based on the most current and relevant data, which helps improve and prolong customer engagement. Let's take a look at the key behaviors of companies that use datagraphs successfully.

They learn at scale and speed.

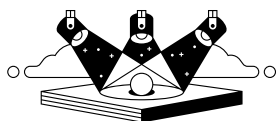
Datagraphs capture how individuals live, work, play, learn, listen, socialize, watch, transact, travel, spend, and do any other activity that can be associated with commerce. Digitalization has made it possible to observe and codify customer data in all these areas

at scale, scope, and speed. Facebook's social graph, for example, analyzes data on 2.8 billion individuals and their social activities from moment to moment: what they're doing, whom they're friending and unfriending, where they're traveling to, what brands they're talking about, what movies they're watching, what music they're listening to, and so on. LinkedIn's professional graph captures in real time how 774 million professionals who work in more than 50 million companies and

attended 90,000-plus educational institutions respond to job postings, status updates, and live videos. Moreover, it maps members to other entities, such as the skills they have, to serve users targeted ads, learning suggestions, news feeds, and more. LinkedIn is now a subsidiary of Microsoft and part of its data ecosystem, which allows it to create an even more vibrant datagraph.

At traditional companies, customer data is stored as independent records in various functional databases. To gain digital advantage, companies must organize data as a graph of interactions that are analyzable by algorithms that provide insight and deliver personalized value to every customer.

They use datagraphs to enrich product offerings. Datagraph leaders organize their knowledge and expertise in machine-readable graph formats with a set of concepts—such as shopping, travel, or search—across categories. Take Airbnb's travel graph. It depicts an inventory of more than 7 million homes, tagged in terms of entities (cities, landmarks, events, and so on), attributes (such as customer reviews and hours of operation), and the relationships among them to yield ever-improving recommendations about not just the type of house to rent but also the best places for dinner or the best times to visit attractions. This ability to expand the product scope allows Airbnb to serve its customers better than traditional hotels, whose data is housed in departmental silos (reservations for the room booking, concierge for restaurant recommendations, spa for massage appointments, and so on). Similarly, Netflix continually improves how it represents and



Spotlight

classifies movies and television shows across 75,000 microgenres (just as Spotify does with music and podcasts).

Google has been able to build something even more powerful. Its Knowledge Graph represents relationships between words and concepts in ways that help its algorithms understand context. This enables Google to respond to verbal queries such as: “Hey, Google, book two tickets to the Colosseum for next Wednesday and charge it to Google Pay.” Because the underlying knowledge is represented as a graph, the algorithms understand what the user is asking; they know that the “Colosseum” is an attraction in Rome, that next Wednesday is May 25, that “book” means to buy tickets, and that “charge” involves using a stored credit card (as opposed to other meanings of these words). And with each query and customer interaction, the Knowledge Graph is refined to reflect new relationships as meanings change.

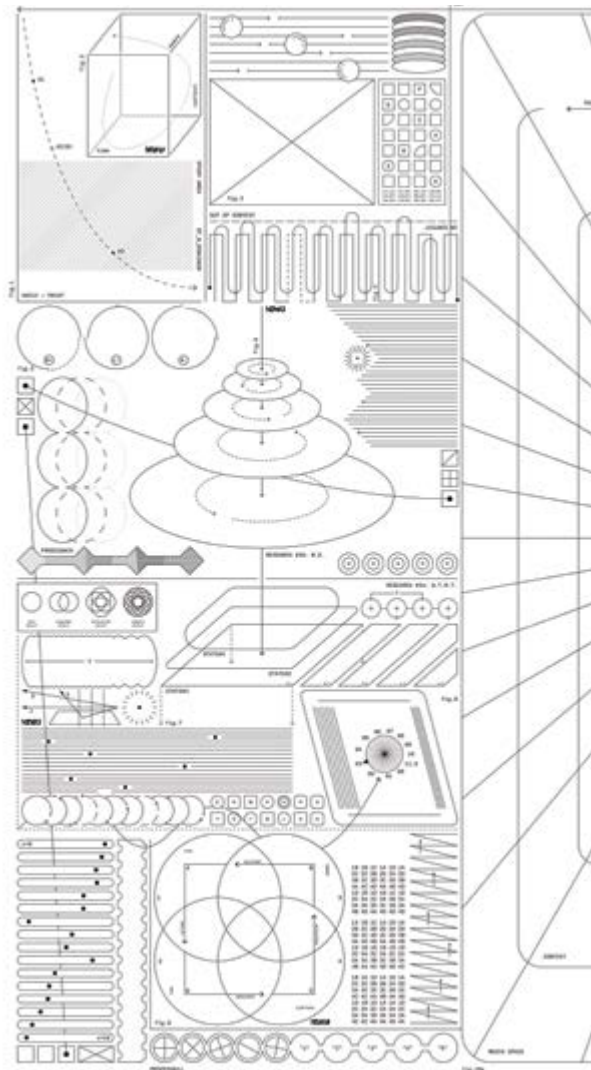
Consider a search query by an avid mountaineer who has hiked Mount Adams and would like to hike Mount Fuji next. She may ask: “What should I do differently to prepare for Mount Fuji compared with Mount Adams?” Today, getting an answer requires multiple searches, but Google is working on a new model with more-complex knowledge linkages (with seamless translation across languages) to more effectively respond to such queries.

To compete with digital giants, ask yourself: Does knowledge about our products exist mostly as separate data sets, or are we developing machine-readable graphs to identify patterns of preference for our customers?

They win customers' moments of truth. In 2001, only 2% of Netflix's recommendations were chosen by its 456,000 users. By 2020, the percentage had increased to 80%, and Netflix had more than 200 million subscribers. Netflix uses its movie graph to win the “moment of truth”: the 90-second-to-two-minute window in which a viewer decides to watch something on Netflix or go elsewhere. Netflix algorithmically customizes and updates its home screen to continuously deliver targeted recommendations for every subscriber. By 2015, Netflix had prevented more than \$1 billion a year in canceled subscriptions thanks to its personalized recommendation engine.

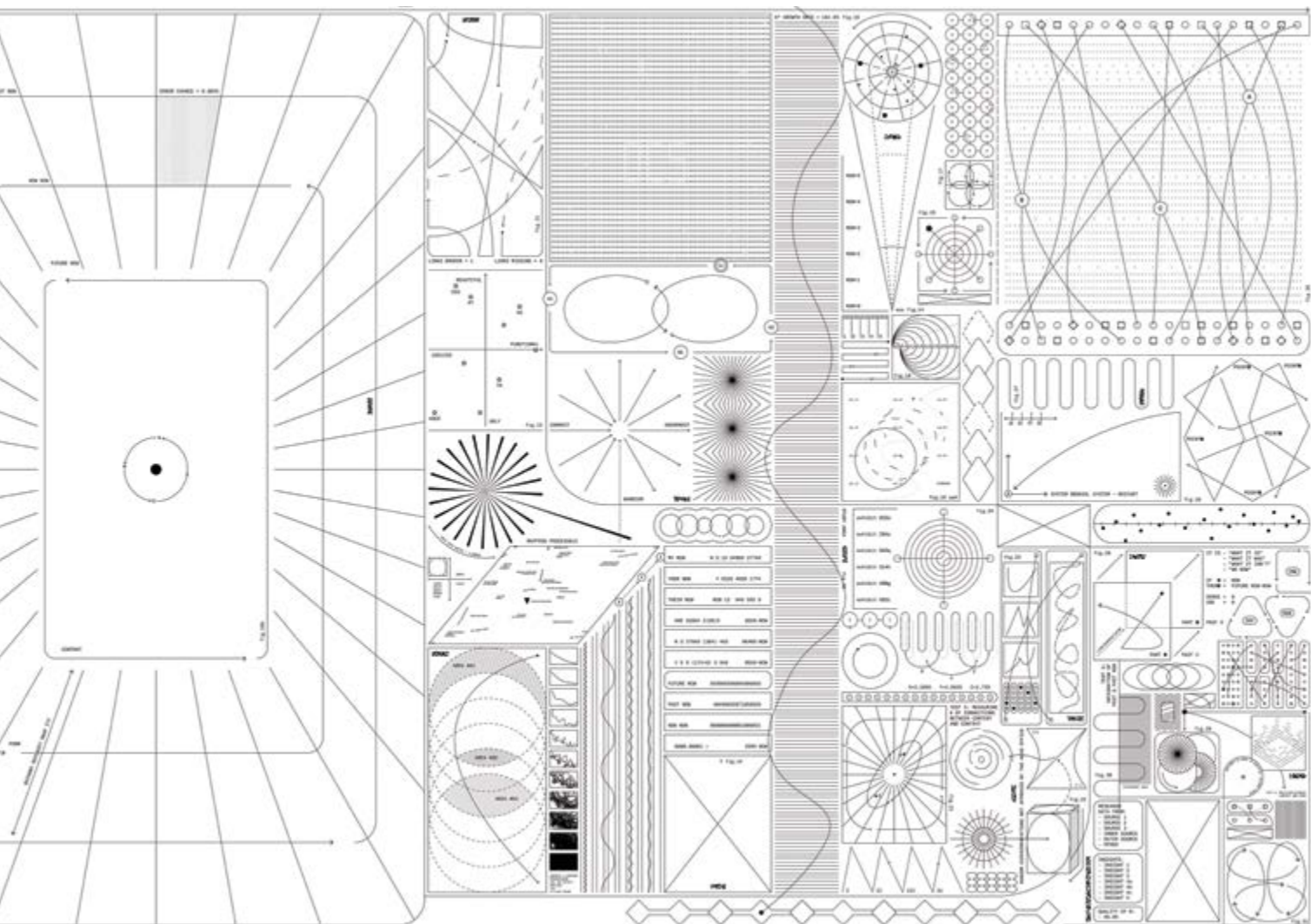
To win its moments of truth, Facebook conducts A/B experiments across 3 billion users in near real time to personalize the social feeds of each user. Before Facebook displays a post, it sorts through an inventory of possibilities and narrows them down to about 500 that past behavior patterns suggest a user is likely to engage with. Then, Facebook's proprietary neural network scores the posts and ranks them before arranging them in a variety of media types, such as text, photos, sounds, and videos interspersed with ads.

Unlike Facebook, whose library of digital content can be instantaneously delivered to its customers worldwide (subject to legal restrictions), Uber's ability to satisfy a customer's need for transportation is based on the availability of vehicles at a precise time and at an exact location. Uber's moment of truth is the five minutes customers are willing to wait for a driver. The ride-sharing company tracks drivers



and passengers who have the app open on their smartphones (it previously tracked users even when they weren't using the app, a controversial policy it was forced to change in 2017 after customer backlash) and uses that data to analyze likely demand patterns. Then it provides incentives for drivers to be available at selected locations. The company continually optimizes its routing algorithms to win customers at the moment of truth.

Although many companies claim to be customer-centric, few use data-graphs and algorithms the way these leaders do. Ask yourself: Are we using



AI-powered algorithms to deliver customers an ever-more-refined product offering to make sure they engage with our product rather than move on?

Getting Started

The first thing businesses that wish to remain competitive against data-graph leaders must understand is that a successful strategy isn't solely dependent on having large volumes of information. It's about collecting relevant product-in-use data in real time to achieve data network effects and build advantage. When businesses

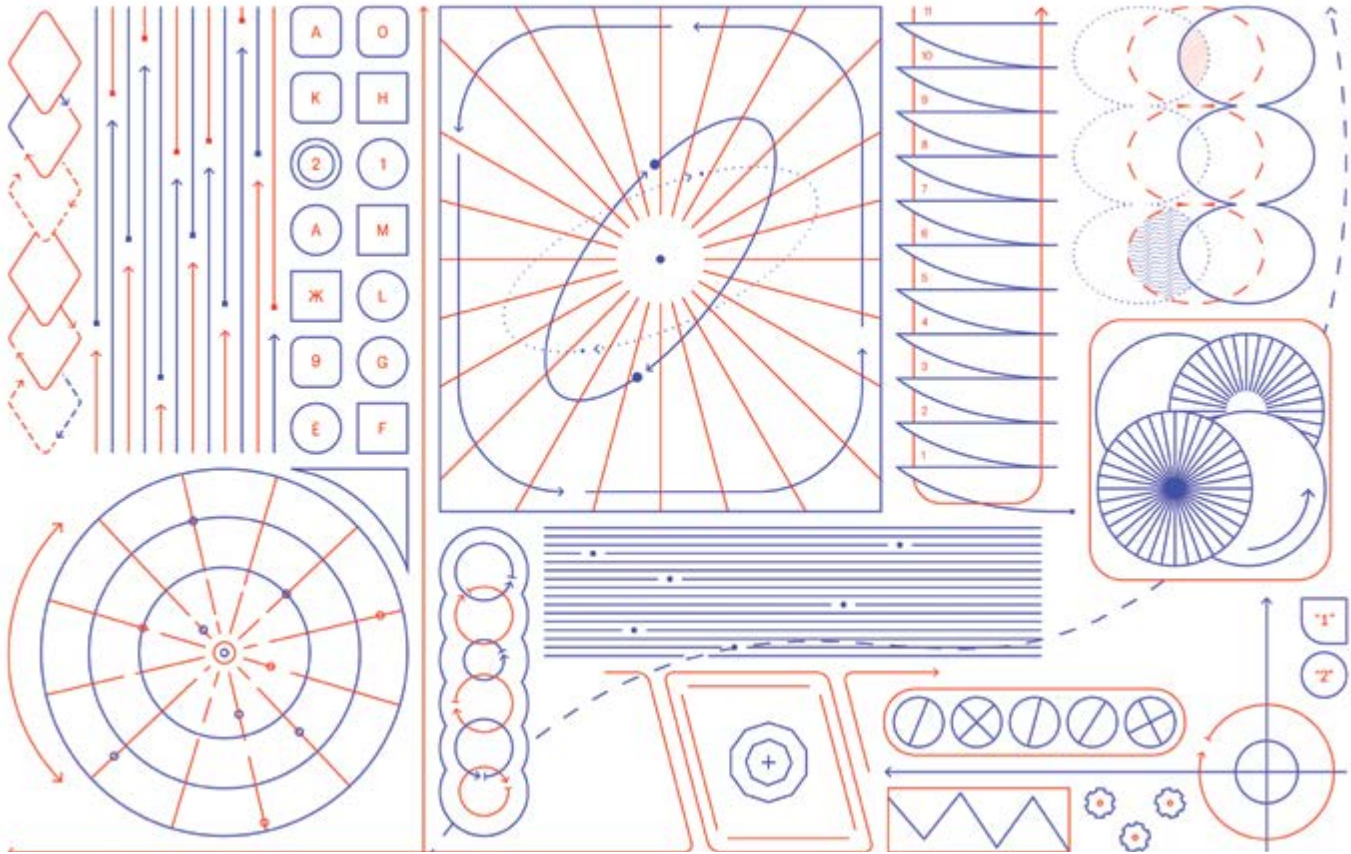
observe more customer interactions with their products, they accumulate richer data; when they sell more products to a more-diverse group of users, they accumulate more-varied data that helps them further differentiate their offerings. Businesses that aren't using datagraphs or have yet to do so successfully must take the following steps to catch up:

1. Develop a datagraph strategy.

To get started, pair executives that have industry knowledge with data scientists to conceptualize your datagraph, examine its future trajectory, and sketch out plausible business implications.

Many companies that don't have the resources of an Amazon or a Netflix have already done this. For example, Stitch Fix was founded as a personalized fashion service in 2010 by a business school student; now, thanks in large part to its fashion graph, its market cap tops \$1.6 billion.

Coursera shows how new entrants can use datagraphs to upend a market. Traditional universities offer "one size fits all" courses and certificates; in contrast, Coursera operates like a Netflix or Amazon for education. It offers a personalized online experience through stackable modules, which are



consumable at different durations, locations, difficulty levels, and price points. It uses its proprietary Skills Graph to customize lifelong learning in ways that traditional universities have not been able to achieve.

Ask yourself how your data offers a unique advantage to your business. You may possess proprietary “data hooks” that allow you to observe at the point of use detailed information that is unavailable to others. Your advantage may come from superior data scope (the depth and richness of your data) and access to complementary data from partners. You may have faster data speed (data in motion compared with a competitor’s episodic data, which is subject to batch processing). Consider how scale, scope, and speed can be increased through acquisitions (consider Microsoft’s acquisitions of

LinkedIn and Activision) or alliances (such as Google’s partnership with Shopify).

2. Develop proprietary algorithms. It’s no longer adequate to carry out different types of analysis independently. Datagraph leaders use proprietary algorithms to conduct descriptive analysis (“What happened?”), diagnostic analysis (“Why did it happen?”), predictive analysis (“What could happen?”), and prescriptive analysis (“What should happen?”) in an overarching framework. You can evolve your datagraph infrastructure from the legacy architectures designed to analyze data at rest (batch processing, independent analysis) to analyze real-time data in motion. Be sure to benchmark your algorithms against others in your industry—and against others of its class. For example, if your success metric is the extent to

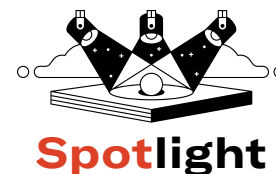
which customers act on your recommendations, how does the performance of your recommendation engine stack up against those of leaders like Netflix, Spotify, and Amazon?

3. Engender trust. Being the custodian of customer data is a huge responsibility. Most customers regard computers, algorithms, and machine learning as complex black boxes, and many believe that their data is being used (even abused) to make digital companies rich and powerful. You must develop ways to use your algorithms to engender trust, and you must earn the right to gather, analyze, and deliver value through data. Explain what you’re doing using language that consumers can understand.

Trust gets eroded when consumers feel that their data is being misused. Facebook has become the poster child



Companies must use algorithms in ways that engender trust, and they must earn the right to gather and analyze data.



for this predicament. Recently, a whistleblower from the Facebook data science team directly accused the company of, among other things, using its data and algorithms to amplify inflammatory content to boost engagement on the site, despite having conducted proprietary internal research showing that doing so would be harmful to users and society. Facebook CEO Mark Zuckerberg has denied the claims: “The argument that we deliberately push content that makes people angry for profit is deeply illogical,” he said. “We make money from ads, and advertisers consistently tell us they don’t want their ads next to harmful or angry content.” Whether Zuckerberg can repair the damage depends on how Facebook uses its algorithms and personal data moving forward and the transparency with which it communicates with users.

Every company must invest resources not only in the technical facets of algorithms but in explaining what they do in ways consumers understand and feel comfortable with. Customers increasingly expect to be informed about how digital products function and AI-supported services are delivered, and countries demand that companies tailor their data operations to local regulations. For example, in Germany, which has strict privacy regulations, Alibaba needs a different data strategy from the one it uses in China. And it must communicate to consumers in both countries in ways that promote trust.

4. Update the organization. Business leaders must allocate the resources necessary to upgrade the technology

infrastructure required for datagraphs. They must recruit talent with breadth and depth in both data science and business. They must structure the data organization as the connective tissue that ties together all parts of the enterprise, recognizing that modern organizations must juggle two powerful, competing factions: those who believe in the supreme power of data and algorithms to solve problems and those who don’t. This tension defines the operating culture of modern organizations: Consider how Netflix CEO Reed Hastings balances the analytical pull of Silicon Valley with the creative pull of Hollywood.

5. Monetize your datagraph.


Datagraphs, when constructed to support and shape strategy, reveal that value lies not only in how products are designed and manufactured but also in how they solve specific problems for customers. Insights from datagraphs will help you choose the most appropriate monetization mechanisms and lay out clear pathways from data to business results. You can defend your current revenue and profits with compelling recommendations based on data network effects, just as Netflix uses real-time data to improve customer retention. You can also use your datagraph to develop more-thoughtful ways to expand your revenue and profit streams by going after new pockets of value, as Apple has done with its foray into credit cards, TV, and health care. And you can counterattack in markets where competitors have already mastered datagraphs, as Disney did with its successful entry into the streaming wars with Disney+.

Reshaping Advantage

We’ve all seen the signs in front of McDonald’s announcing, “Over X Billion Served” and have watched the number rise over the years. But tracking how many burgers are sold every year is a relic of the past. Datagraph leaders care less about absolute numbers. Instead, they ask: Do we have data on where each consumer buys her burgers? At what time? What does she drink with it? What does she do before or after buying a burger? Who are our customers and what are their ages, income, location, preferences, lifestyles, and so on? How can we satisfy more of their needs so that they spend more dollars with us than with someone else, feel confident that they got value for their money, and keep coming back?

Datagraphs will reshape competition in every sector sooner than most expect. It’s time for every company to move beyond using data to improve operational efficiency and recognize the competitive advantage of datagraphs. Senior leaders must invest in upgrading their data architecture to enable a real-time, comprehensive view of how consumers interact with their products and services. With this structure in place, leaders can develop unique ways to solve customer problems. ☺ **HBR Reprint S22033**

Spotlight Package Reprint R2203B

 **VIJAY GOVINDARAJAN** is the Coxie Distinguished Professor at Dartmouth’s Tuck School of Business and an executive fellow at Harvard Business School. **N. VENKAT VENKATRAMAN** is the David J. McGrath Jr. Professor in Management at Boston University’s Questrom School of Business.