The Destiny of Silicon Machine

Sushil Singh
B.Tech CSE, Final Year
Lingaya's GVKS Institute of Management and Technology
Old Faridabad, Haryana 110062
Email: -sushiljadouns@gmail.com

Abstract—Computer science is undergoing a fundamental change that is changing our mindset towards the world. An important aspect of this change is the theory and applicationsdealing with the gathering and analysing of large real-world data sets. In this paper, weintroduce four research projects in which processing and interpreting large data sets is a central focus. Innovative ways of analysing such data sets allow us to extract useful informationthat we would never have obtained from small or synthetic data sets, thus providing us with new insights into the modern world.

Keywords—Modern computer science; social networking websites; large data bases; high- complesx data.

1.Introduction

Modern computer science is undergoing a fundamental change. In the early years of the field, computer scientists were primarily concerned with the size, efficiencyand reliability of computers. They attempted to increase the computational speedas well as reduce the physical size of computers, to make them more practical anduseful. The research mainly dealt with hardware, programming languages, compilers,operating systems and data bases. Meanwhile, theoretical computer science developedan underlying mathematical foundation to support this research which in turn, ledto the creation of automata theory, formal languages, computability and algorithmanalysis. Through the efforts of these researchers, computers have shrunk from thesize of a room to that of a dime, nearly every modern household has access to theinternet and communications across the globe are virtually instantaneous.
Computers can be found everywhere, from satellites hundreds of miles above us to pacemakers inside eating human hearts. The prevalence of computers, together with communication devices and data storage devices, has made vast quantities of data accessible. This data incorporates important information that reveals a closer approximation of the real world and is fundamentally different from what can be extracted from individual entities. Rather than analysing and interpreting individual messages,

.

2.Innovative Research Projects

Traditional research in theoretical computer science has focused primarily on problems with small input sets. For instance, in the past, stores tracked the items purchased by each individual customer and gave that customer discounts for future purchases of those items. However, with the help of modern algorithms, service providers such as Netflix are now able to, not only make predictions based on a customer's past preferences, but amalgamate preferences from millions of customers to make accurate and intelligent suggestions and effectively increase sales revenue. The following subsections describe four ongoing research projects involving the analysis and interpretation of large data sets. Each represents a promising direction for rediscovering fundamental properties of large-scale networks that will reshape our understanding of the world.
Computer science needs more people and more diversity. The modern world relies on information technologies, but is producing too few computer scientists to meet its needs. Currently, most computer scientists are White and Asian males, which is one of the shrinking demographics in society. Using more women and under-represented minorities, more jobs could be filled and a more diverse range of perspectives would increase design creativity and innovation in computing technology.

2.1Tracking communities in social networks

A social network is usually modeled as a graph in whichvertices represent entitiesand edges represent interactions between pairs of entities. In previous studies, a community was often defined to be a subset of vertices that are densely connectedinternally but sparsely connected to the rest of the network. Accordingly, the bestcommunity of the graph was typically a peripheral set of vertices barely connected tothe rest of the network by a small number of edges. However, it is our view that forlarge-scale real-world societies, communities, though better connected internally thanexpected solely by chance, may also be well connected to the rest of the network.
It is hard to imagine a small close-knit community with only a few edges connecting it to the outside world. Rather, members of a community, such as a computer science department, are likely to have many connections outside the community, such as family, religious groups, other academic departments and so on. Empirically, a community displays a higher than average edge to vertex squared ratio which reflects the probability of an edge between two randomly-picked vertices,

but can also be connected to the rest of the network by a significant number of edges, which may even be larger than the number of its internal edges.

With this intuitive notion of community, two types of structures are defined: the “whiskers” and the “core”. Whiskers are peripheral subsets of vertices that arebarely connected to the rest of the network, while the core is the central piece thatexclusively contains the type of community we are interested in. Then, the algorithmfor finding a community can be reduced to two steps: 1) identifying the core in whichno whiskers exist, and 2) identifying communities in the core. Further, extractingthe exact core from both weighted and unweighted graphs has been proved to beNP-complete. Alternative heuristic algorithms have been developed, all of which arecapable of finding an approximate core, and their performance can be justified by theexperimental results based on various large-scale social graphs. In this way, onecan obtain communities that are not only more densely connected than expected bychance alone, but also well connected to the rest of the network.Much of the early work on finding communities in social networks focused on partitioning the corresponding graph into disjoint subcomponents. Algorithmsoften required dense graphs and conductance was widely taken as the measure of thequality of a community. Given a graph $G = (V; E)$, the *conductance* of a subset of vertices $S\ \mu\ V$ is defined as:

$$\varphi(S) = \frac{\sum_{i\in S, j\notin S} A_{ij}}{\min\left\{\sum_{i\in S, j\in V} A_{ij},\ \sum_{i\notin S, j\in V} A_{ij}\right\}},$$

where $fA_{ij}ji;\ j\ 2\ V\ g$ are the entries of the adjacency matrix for The graph. A subsetwas considered to be community-like when it had low conductance value. However, asdiscussed earlier, an individual may belong to multiple communities at the same timeand will likely have more connections to individuals outside of his/her communitythan inside. One approach to finding such well-defined overlapping communities isthat of Mishra et al.where the concept of an (α, β)-community was introduced andseveral algorithms were given for finding an (α, β)-community in a dense graph under certain conditions. Given a graph $G = (V; E)$ with self-loops added to all vertices, asubset $C\ V$ is called an (α, β)-*community* when each vertex in C is connected toat least βvertices of C (self-loop counted) and each vertex outside of C is connected to at most α vertices of C ($\alpha<\beta$).

Among the interesting questions being explored are why (α, β)-communities correspond to well-defined clusters and why there is no sequence of intermediate (α, β)-communities connecting one cluster to another. Other intriguing questions includewhether different types of social networks incorporate fundamentally different socialstructures; what it is about the structure of real-world social networks that leads tothe structure of cores, as in the Twitter graph, and why some synthetic networks donot display this structure.

By taking the intersection of a group of massively overlapping (α, β)-communitiesobtained from repeated experiments, one can eliminate random factors and extractthe underlying structure. In social graphs, for large community size k, the (α, β)-communities are well clustered into a small number of disjoint cores, and there areno isolated (α, β)-communities scattered between these densely-clustered cores. Thenumber of cores decreases askincreases and becomes relatively small for large k. Thecores obtained for a smaller k either disappear or merge into the cores obtained fora larger k. Moreover, the cores correspond to dense regions of the graph and thereare no bridges of intermediate (α, β)-communities connecting one core to another.

In contrast, the cores found in several random graph models usually have significantoverlap among them, and the number of cores does not necessarily decrease askincreases. Extensive experiments demonstrate that the core structure displayed invarious large-scale social graphs is, indeed, due to the underlying social structureof the networks, rather than due to high-degree vertices or to a particular degreedistribution.

This work opens up several new questions about the structure of large-scale socialnetworks, and it demonstrates the successful use of the (α, β)-community algorithmon real-world networks for identifying their underlying social structure. Further, this work inspires an effective way of finding overlapping communities and discovering theunderlying core structure from random perturbations. In social graphs, one wouldnot expect a community to have an exact boundary; thus, the vertices inside an(α, β)-community but outside the corresponding core are actually located in the roughboundary regions. Other open questions include how the core structure will evolve,whether the cores correspond to the stable backbones of the network, and whether thevertices that belong to multiple communities at the same time constitute the unstableboundary regions of the network.

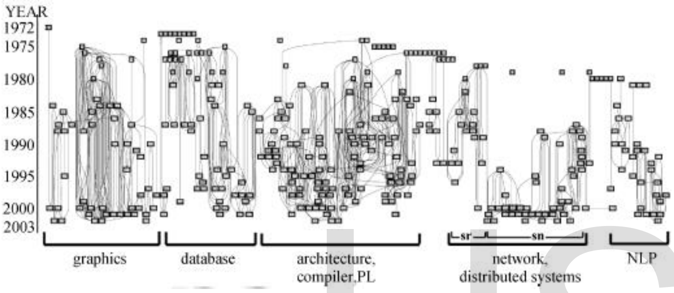
2.2 Tracking flow of ideas in scientific literature

Such information-seeking activities often require the ability to identify topics withtheir time of appearance and to follow their evolution. Recently, in their unpublishedwork, Jo et al. have developed a unique approach to achieving this goal in a time-stamped document collection with an underlying document network which representsa wide range of digital texts available over the internet. Examples are scientific papercollections, text collections associated with social networks such as blogs and Twitter,and more generally, web documents with hyperlinks. A document collection withoutan explicit network can be converted into this format by connecting textually similardocuments to generate a document network.

The approach emphasizes discovering the topology of topic evolution inherent ina corpus. As demonstrated in the above work, the topology inherent in the corpuscarries surprisingly rich information about the evolution of topics. Topics, alongwith the time that they start to appear in the corpus, can

International Journal of Scientific & Engineering Research Volume 7, Issue 12, December-2016
ISSN 2229-5518

be identified by visitingeach document in the corpus chronologically and determining if it initiates a topic.
A document is considered to initiate a topic if it has a textual content that is notexplained by previously discovered topics and persists in a significant number of laterdocuments. After topics are obtained by the chronological scan, an associated graphcan be built whose vertices are topics and whose edges reflect cross-citation relationsbetween topics. Globally, this generates a rich topological map showing the landscapeof topics over time. Figure 3 shows the results of the work by Jo et al. applying thisapproach to the ACM corpus. Topics in the network research area emerged in the1980s without significant ancestors, while the areas of compiler and graphics researchexhibit steady evolution with an appreciable number of topics in the early years of theACM corpus. We can also construct an individual topic evolution graph for a givenseed topic, and such a graph may contain multiple threads indicating that the seedtopic has been influenced by multiple fields. The relationship between these threadsmay change over time as well.



Related to this research, content-based inferring and learning has been extensively studied recently. Various methods to improve question-answer services in social networks have been proposed. In addition, tracking popular events in socialcommunities can be achieved using a statistical model.

2.3Reconstructing networks

The study of large networks has brought about many interesting questions, suchas how to determine which members of a population to vaccinate in order to slow thespread of an infection, or where to place a limited number of sensors to detect the flowof a toxin through a water network. Most algorithms for solving such questions makethe assumption that the structure of the underlying network is known. For example,detectives may want to use such algorithms to identify the leaders of a criminalnetwork, and to decide which members to turn into informants. Unfortunately, theexact structure of the criminal network cannot be easily determined. However, itis possible that the police department has some information about the spread of acertain property through the network; for instance, some new drug may have firstappeared in one neighbourhood, and then in two other neighbourhoods, and so on. Thework by Soundarajan et al. attempts to create algorithms to recover the structureof a network given information about how some property, such as disease or crime,has spread through the network.
This work begins by defining a model of contagion describing how some propertyhas spread through a network. The model of contagion for information spread may be: “a vertex learns a piece of information in the time interval after one of its neighbourslearns that information.” A more complex model of contagion corresponding to thespread of belief may be: “a vertex adopts a new belief in the time interval after aproportion *p* of its neighbours adopts that belief.” For example, a person will probablynot join a political party as soon as one of his friends joins that party, but he mayjoin it after two-thirds of his friends have joined it.Next, the network recovery algorithm assumes that vertices are partitioned intodiscrete time intervals, corresponding to the time when they adopt the property.
For a given model of contagion, the algorithm attempts to find a network over theset of vertices such that when the property in question (e.g. information, belief)is introduced to some vertices in the first time interval, and then spreads to othervertices in accordance with the model of contagion, every vertex adopts the propertyat an appropriate time. Initial work has created such algorithms for two models ofcontagion: the model corresponding to the spread of information, where a vertexadopts a property in the time interval after one of its neighbours has adopted thatproperty, and the model corresponding to the spread of belief, where a vertex adoptsa property in the time interval after at least half of its neighbours have adopted thatproperty.Future work will focus on finding algorithms for other models of contagion, especially the models in which a vertex adopts the property after a proportion *p* ofits neighbours has adopted that property, for arbitrary values of *p*. Other directionsinclude finding algorithms for networks in which there are two or more propertiesspreading through the network. This work also opens up questions about the typesof graphs produced by these algorithms. For instance, do all possible graphs havesome edges in common? Are there any edges that do not appear in any of the solution graphs? Which edges are the most or least likely? Related research in this area includes work by Gomez-Rodriguez et al, which studied information flow and cascades in online blogs and news stories. Work by Leskovec et al. studied the question of how to detect outbreaks or infection in a network. In addition, a more general problem of link prediction was studied by Clauset et al.

2.4Tracking bird migration in north America

Hidden Markov models (HMMs) assume a generative process for sequential datawhereby a sequence of states (i.e. a sample path) is drawn from a Markov chain in ahidden experiment. Each state generates an output symbol from a given alphabet, andthese output symbols constitute the sequential data (i.e. observations). The classicsingle path problem, solved by the Viterbi algorithm, is to find the most probablesample path given certain observations for a given Markov model.

Two generalizations of the single path problem for performing collective inferenceon Markov models are introduced in Ref., motivated by an effort to model birdmigration patterns using a large database of static observations. The eBird databasemaintained by the Cornell Lab of Ornithology contains millions of bird observationsfrom throughout North America reported by the general public using the eBird webapplication. Recorded observations include location, date, species and number ofbirds observed. The eBird data set is very rich, and the human eye can easily discernmigration patterns from animations showing the observations as they unfold overtime on a map of North America. However, the eBird data entries are static andmovement is not explicitly recorded, only the distributions at different points in time.Conclusions about migration patterns are made by the human observer, and the goalis to build a mathematical framework to infer dynamic migration models from thestatic eBird data. Quantitative migration models are of great scientific and practicalimportance. For example, this problem comes from an interdisciplinary project atCornell University to model the possible spread of avian influenza in North Americathrough wild bird migration.

The migratory behaviour of a species of birds can be modelled by a single generative process that independently governs how individual birds by between locations.This gives rise to the following inference problem: a hidden experiment draws manyindependent sample paths simultaneously from a Markov chain, and the observationsreveal collective information about the set of sample paths at each time step, fromwhich the observer attempts to reconstruct the paths.

Displays the pattern of ruby-throated hummingbird migration inferredby this model for the four weeks starting on the dates indicated. The top row shows thedistributions and migrating paths inferred by the model: grid cells coloured in lightershades represent more birds; arrows indicate fight paths between the week shownand the following week, with line width proportional to bird flow. The bottom rowshows the raw data for comparison: white dots indicate negative observations; blacksquares indicate positive observations, with size proportional to bird count; locationswith both positive and negative observations appear a charcoal colour. This leads toa somewhat surprising prediction that when migrating north, some hummingbirds will be across the Gulf of Mexico while others follow the coastline, but when flying
south, they generally stay above land. This prediction has been confirmed by work
performed by ornithologists. For example, in the summary paragraph on migrationfrom the Archilochus colubris species account, Robinson et al. write “Many byacross Gulf of Mexico, but many also follow coastal route. Routes may differ fornorth- and southbound birds." The inferred distributions and paths are consistentwith both seasonal ranges and written accounts of migration routes.

3 Theoretical Foundation

As demonstrated in the previous section, the focus of modern computer scienceresearch is shifting to problems concerning large data sets. Thus, a theoretical foundation and science base is required for rigorously conducting studies in many relatedareas. The theory of large data sets is quite different from that of smaller data sets;when dealing with smaller data sets, discrete mathematics is widely used, but forlarge data sets, asymptotic analysis and probabilistic methods must be applied. Additionally, this change in the theoretical foundation requires a completely differentkind of mathematical intuition.

3.1Large-Scale graphs

Large graphs have become an increasingly important tool for representing real-world data in modern computer science research. Many empirical experiments havebeen performed on large-scale graphs to reveal interesting findings. Acomputer network may have consisted of only a few hundred nodes in previous years,but now we must be able to deal with large-scale networks containing millions oreven billions of nodes. Many important features of such large graphs remain constantwhen small changes are made to the network. Since the exact structure of largegraphs is often unknown, one way to study these networks is to consider generativegraph models instead, where a graph is constructed by adding vertices and edges ineach time interval. Although such graphs typically differ from real-world networks inmany important ways, researchers can use the similarities between the two types ofnetworks to gain insight into real-world data sets.

A simple but commonly used model for creating randomgraphs is the ErdÄos-Renyi model, in which an edge exists between each pair of vertices with equal probability, independent of the other edges. A more realistic model is known as the “preferential attachment” model, in which the probability that an edge is adjacent toa particular vertex is proportional to the number of edges already adjacent to thatvertex. In other words, a high degree vertex is likely to gain more edges than a lowdegree vertex. The preferential attachment model gives rise to the power-law degreedistribution observed in many real-world graphs.

Another interesting feature of real-world networks is the existence of the “giantcomponent". The following table describes the number of components of each size ina protein database containing 2,730 proteins and 3,602 interactions between proteins.

Component Size	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	···	1000
# Components	48	179	50	25	14	6	4	6	1	1	1	0	0	0	0	1	···	0

As the component size increases, the number of components of that size decreases.Thus, there are many fewer components of size four or more in this protein graph thancomponents of size one, two, or three. However, those smaller components containonly 899 proteins, while the other 1,851 proteins are all contained within one giantcomponent of size 1,851, as shown in the following table.

Component Size	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	···	1851
# Components	48	179	50	25	14	6	4	6	1	1	1	0	0	0	0	1	···	1

Consider the ErdÄos-Renyi random graph model in which each edge is addedindependently with equal probability. Suppose that we start with 1,000 vertices andzero edges. Then, there are clearly 1,000 components of size one. If we add one edge,we will have 998 components of size one and one component of size two. However, agiant component begins to emerge as more edges are added, as shown in the followingtable. The graph contains many components of small size and a giant component ofsize 101. This occurs because a component is more likely to attract additional verticesas its size increases.

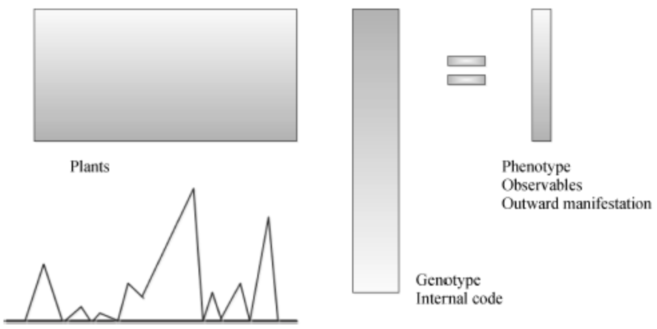
Size of Component	1	2	3	4	8	10	14	20	55	101
Number of Components	367	70	24	12	2	2	2	1	1	1

Since random graph models mimic some vital features of real-world networks,it is often helpful to study the processes that generate these features in randomgraph models. An understanding of these processes can provide valuable insights foranalyzing real-world data sets.

3.2Sparse vectors

Having sketched an outline of a science base for high-dimensional data, we nowfocus on studying sparse vectors as a science base for a number of application areas.Sparse vectors are useful to reduce the time required to find an optimal solution andto facilitate reconstructions and comparisons. For example, plant geneticistsare interested in determining the genes responsible for certain observable phenomena.The internal genetic code is called the genotype and the observable phenomena oroutward manifestation is called the phenotype. Given the genotype for a numberof plants and some output parameter, one would attempt to determine the genesresponsible for that particular phenotype.

Solutions to this type of problem are generally very sparse. Intuitively, one wouldexpect this since only a small set of genes is probably responsible for the observedmanifestation. This situation arises in a number of areas and suggests the followingunderlying problem: given a matrix A, a sparse vector x and a vector b where $Ax= b$,how do we ¯nd the sparse vector x knowing A and b? This problem can be formallywritten as:



minimize $||x||$subject to $Ax= b$

where the 0-norm is the number of non-zero elements of the sparse vector x. However, the 0-norm is non-convex and optimization problems of this nature are often NP-complete. Thus, the question remains of when the solution to the convex optimization problem:

minimize $||xj||$subject to $Ax= b$

correctly recovers the vector x. Note that the above minimization problem can be solved by linear programming.

4.Moores Law

Moore's lawis the tells that the number of transistors in a dense integrated circuit doubles approximately every two years. The law is named after Gordon Moore, the co-founder of Fairchild Semiconductor and Intel, whose 1965 paper described a doubling every year in the number of components per integrated circuit, and projected this rate of growth would continue for at least another decade. In 1975, looking forward to the next decade, he revised the forecast to doubling every two years. The period is often quoted as 18 months because of Intel executive David House, who predicted that chip performance would double every 18 months (being a combination of the effect of more transistors and the transistors being faster). Moore's prediction proved accurate for several decades, and has been used in the semiconductor industry to guide long-term planning and to set targets for research and development. Advancements in digital electronics are strongly linked to Moore's law: quality-adjusted microprocessor pricess, memory capacity, sensors and even the number and size of pixels in digital cameras. Digital electronics has contributed to world economic growth in the late twentieth and early twenty-first centuries. Moore's law describes a driving force of technological and social change, productivity, and economic growth.

Moore's law is an observation or projection and not a physical or natural law. Although the rate held steady from 1975 until around 2012, the rate was faster during the first decade. In general, it is not logically sound to extrapolate from the historical growth rate into the indefinite future. For example, the 2010 update to the International Technology Roadmap for Semiconductors, predicted that growth would slow around 2013, and Gordon Moore in 2015 foresaw that the rate of progress would reach saturation: "I see Moore’s law dying here in the next decade or so." Intel stated in 2015 that the pace of advancement has slowed, starting at the 22 nm feature width around 2012, and continuing at 14 nm. Brian Krzanich, CEO of Intel, announced that "our cadence today is closer to two and a half years than two.” This is scheduled to hold through the 10 nm width in late 2017.[20] He cited Moore's 1975 revision as a precedent for the current deceleration, which results from technical challenges and is “a natural part of the history of Moore's law.”

Today hardware has to be designed in a multi-core manner to keep up with Moore's law. In turn, this also means that

IJSER

software has to be written in a multi-threaded manner to take full advantage of the hardware

5 . Neural Impulse Actuater

The Neural Impulse Actuator (NIA**)** is a brain–computer interface (BCI) device developed by OCZ Technology. BCI devices attempt to move away from the classic input devices like keyboard and mouse and instead read electrical activity from the head, preferably the EEG. The name Neural Impulse Actuator implies that the signals originate from some neuronal activity; however, what is actually captured is a mixture of muscle, skin and nerve activity including sympathetic and parasympathetic components that have to be summarized as biopotentials rather than pure neural signals. As of May 27, 2011, the OCZ website says that the NIA is no longer being manufactured and has been end-of-lifed.

The name Neural Impulse Actuator is still justifiable since also the secondary signals are under neuronal control. The biopotentials are decompiled into different frequency spectra to allow the separation into different groups of electrical signals. Individual signals that are isolated comprise alpha and beta brain waves, electromyograms and electro oculograms. The current version of the NIA uses carbon-fibers injected into soft plastic as substrate for the headband and for the sensors and achieves sensitivity much greater than the original silver chloride-based sensors using a clip-on interface to the wire harness.

Control over the computer in either desktop or gaming environments is done by binding keys to different zones within as many as three vertical joysticks. Each joystick can be divided into several zones based on thresholds and each zone within each joystick can be bound to a keyboard key. Each keystroke can further be assigned to several modes, including single keystroke, hold, repeat and dwell, which allows full plasticity with respect to configuration of the NIA for any application. Moreover, the same "vertical joysticks" can be used in more than one instance to enable simultaneous pressing of multiple keys at any given time like "W" and "Spacebar" for jumping forward or toggling between left and right strafing for running in a zigzag pattern.

6. ARPANET TO LIFI

Surfing the Internet, creating and finding hosting for a Web site, and sharing with social media are all things we take for granted today—but none of them would be possible without the efforts of the innovators who established what we now know as the modern Web.

In October of 1969, four leading U.S. universities—University of California Los Angeles (UCLA), Stanford Research Institute (SRI), University of California Santa Barbara (UCSB), and the University of Utah—activated a project known as the Advanced Research Projects Agency Network (ARPANET), creating the first successful network of computers in a time when computers barely interacted with their users, let alone one another. This historic connection provided the basis for the Internet as we know it today.

Over the nearly five decades since ARPANET's debut, computer networking has evolved to a level beyond even the loftiest dreams of those who created it. The next time you shoot a quick email to a friend and send it winging through the electronic ether, consider this: The first message ever sent on a network consisted of just two letters. The message was meant to read, "LOGIN," but the network only managed to transmit two letters before the whole system crashed. Not a very auspicious beginning for a system that would one day support 297 billion emails every single day.

Even the enormous amount of data created by email is but an eddy in the raging river of data that surges through the modern Internet. But it wasn't always so. In 1984, when ARPANET was released from military control and began to merge with the National Science Foundation Network (NSFNET) to form what we now call "the Internet," the cutting-edge hardware that carried its traffic pushed data at 56 Kilobytes(K) per second. That's a speed best remembered as the fastest possible in the not-so-distant days before broadband Internet. By way of comparison, the average Internet access speed in the U.S. today is 7.6 Megabytes (MB) per second—roughly 136 times faster.

And Netizens—with our online shopping, hosting for personal and professional websites, email and social media—use every bit of that extra bandwidth. In just one minute on an average day, 700 videos, 28,000 Tumblr posts, 100,000 tweets and more than 34,000 Facebook "likes" hit the 'net, and with total Internet traffic expected to quadruple by 2014, our "need for speed" is only likely to grow**.**

Haas, a professor of mobile communications at the University of Edinburgh in Scotland, has been championing the idea that data can be transmitted through LED lightbulbs for years. Now, he has created a working model of a "Li-Fi" system.

In a recent TED talk, Haas demonstrated one of these Li-Fi prototypes, transmitting a video from a store-bought LED lamp to a solar cell to a laptop.

“Li-Fi is essentially the same as Wi-Fi, except for a small difference—we use LED lights around us to transmit the data wirelessly as opposed to using radio,” Haas says.

Traditional Wi-Fi uses radio signals to transmit data to devices, such as phones and laptops. Currently, Wi-Fi carries about half of the world’s internet transmissions. This percentage is expected to grow in coming years as more people get online and as the “Internet of Things” (objects with internet connectivity, from remotely programmable coffee makers to smart cars) expands. Some experts, including Haas, worry that this will create a so-called “spectrum crunch,” where Wi-Fi networks slow under heavy demand.

“Radio spectrum is not sufficient,” Haas says. “It’s heavily used, it’s very crowded…we see that when we go to airports and hotels, where many people want to access the mobile internet and it’s terribly slow. I saw this coming 12, 15 years ago, so I thought ‘what are better ways of transmitting data wirelessly?’”

The idea of transmitting data through the visible light spectrum is not new. Alexander Graham Bell transmitted sound via a beam of sunlight in 1880 using a photophone, a sort of solar-powered wireless telephone. In the past several decades, a number of researchers have looked at using visible light to transmit data.

But what Haas seized on—the key to Li-Fi—is the use of simple LED lightbulbs for data transmission. When Haas first started looking at alternative wireless systems, LED bulbs were becoming more widespread in homes, thanks to their energy savings over traditional incandescent bulbs. LED bulbs are controlled by a driver, which can rapidly dim the light or turn it on or off. Therefore, Haas figured, data could be encoded in subtle shifts of the light’s brightness, shifts imperceptible to the human eye.

So Haas and his students began to experiment with an IKEA lamp, replacing its incandescent bulb with an LED bulb. Eventually, they created a working transmitter and receiver system with the lamp and a solar panel. Fittingly, their research was done in a University of Edinburgh building named after Alexander Graham Bell, who was born in Scotland.

Li-Fi stands to be much faster than Wi-Fi. In recent experiments, researchers have been able to reach Li-Fi speeds as fast as 224 gigabits per second. At these speeds, a person could download nearly 20 full-length movies in a single second. According to Haas's research, Li-Fi can achieve data density 1,000 times greater than Wi-Fi, because Li-Fi signals are contained in a small area, as opposed to the more diffuse radio signals.

In addition to being faster than Wi-Fi, Li-Fi will be more secure, Haas says. While Wi-Fi signals can pass through walls (allowing your neighbors to “share” your connection), home Li-Fi signals can be kept indoors by drawing the curtains. The system wouldn’t mean having to keep your lights on all the time either, Haas says—bulbs could be dimmed to such a point that they appear off, but still transmit data.

Now, Haas’ company, pureLiFi, has begun mass producing Li-Fi routers for a limited corporate clientele. They hope to bring them to a widespread market in the next several years. Li-Fi could make its way into business and industrial uses in the next two years or so. From there, it might not be long until it finds its way into homes. The system can easily network any device with an LED light—an electric kettle, an oven. Ultimately, this could bring about the Internet of Things era much faster. Haas also sees Li-Fi as a way to bring internet to remote locations, using hilltop transmitters and rooftop solar panels. LED streetlights could even be used to form a network of outdoor Li-Fi, making it possible to stay connected when walking around the city.

Just how quickly Li-Fi could spread remains unclear. "Li-Fi technology thus offers numerous benefits but there are certa in barriers that must be overcome before it becomes a ubiquitous part of our lives," write researchers from St. Xavier's College in Kolkata, India. These barriers include the fact that Li-Fi becomes less powerful when light is blocked, whether due to fog or other conditions. Nevertheless, the paper says, the Li-Fi industry is slated to be worth \$6 billion by 2018.

Haas and his team aren’t the only people experimenting with Li-Fi. Chinese researchers have developed a basic Li-Fi prototype as well, powering several laptops with one LED bulb. The Fraunhofer Institute, a German research organization, has been working on Li-Fi hotspot prototypes as well. Even NASA recently announced plans to study Li-Fi’s potential uses in space travel.

IJSER

“The incandescent lightbulb delivers illumination,” Haas says. “In 20 years, the [LED] lightbulb will deliver hundreds of applications."

7.Conclusion

Future computer science research is believed to employ, analyze, and interpretlarge data sets. In this paper, we have discussed several examples of current projectsthat represent modern research directions in computer science, ranging from identifying communities in large-scale social networks to tracing bird migration routes inNorth America. As computing pervades every facet of our lives and data collectionbecomes increasingly ubiquitous, feasible algorithms for solving these problems arebecoming more and more necessary to analyze and understand the vast quantities ofavailable information. In order to rigorously develop these algorithms, a mathematical foundation must be established for large data sets, including the theory of largegraphs, high-dimensional data, sparse vectors and so on. These innovative studiesdiscover striking results that reveal a fundamental change in computer science thatwill reshape our knowledge of the world.

Computer science is key to solving the world’s most crucial problems -- environmental sustainability, poverty, hunger and homeland security. It can give the power to shape one’s future and possibly change the world, solvingissues that face the world today or invent a great innovation to make the world a better place. Computer science involves problem solving, computational thinking and abstract reasoning.

References

- Shaparenko B, Caruana R, Gehrke J, Joachims T. Identifying temporal patterns and key playersin document collections. Proc. of IEEE ICDM Workshop on Temporal Data Mining: Algorithms, Theory and Applications (TDM-05). 2005. 165-174.
- Gaertler M. Clustering. Network Analysis: Methodological foundations, 2005, 3418: 178-215.
- Leskovec J, Lang K, Dasgupta A, Mahoney M. Statistical properties of community structure inlarge social and information networks. Proc. of 18th International World Wide Web Conference
- Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Phys. Rev.E, 2004, 69(026113).
- Wang L, Hopcroft JE. Community structure in large complex networks. Proc. of 7th Annual Conference on Theory and Applications of Models of Computation (TAMC). 2010.
- Clauset A, Newman MEJ, Moore C. Finding community structure in very large networks. Phys.Rev. E, 2004, 70(06111).
- Girvan M, Newman MEJ. Community structure in social and biological networks. Proc. OfNatl. Acad. Sci. USA. 2002, 99(12): 7821-7826.
- Newman MEJ. Detecting community structure in networks. The European Physical J. B, 2004,38: 321-330.
- Newman MEJ. Fast algorithm for detecting community structure in networks. Phys. Rev. E,2004, 69(066133).
- Newman MEJ. Finding community structure in networks using the eigenvectors of matrices.Phys. Rev. E, 2006, 74(036104).
- Newman MEJ. Modularity and community structure in networks. Proc. of Natl. Acad. Sci.USA. 2006, 103(23): 8577-8582.
- Lang K, Rao S. A °ow-based method for improving the expansion or conductance of graph cuts. Proc. of 10th International Conference Integer Programming and Combinatorial Optimization(IPCO). 2004.
- Schae®er SE. Graph clustering. Computer Science Review, 2007,1(1): 27-64.
- Mishra N, Schreiber R, Stanton I, Tarjan RE. Finding strongly-knit clusters in social networks.Internet Mathematics, 2009, 5(1-2): 155-174.
- He J, Hopcroft JE, Liang H, Supasorn S, Wang L. Detecting the structure of social networksusing (α,β)-communities. Proc. of 8th Workshop on Algorithms and Models for the Web Graph(WAW). 2011.

IJSER