
Social Semantic Information Spaces

John G. Breslin¹

Digital Enterprise Research Institute, National University of Ireland, Galway*
firstname.lastname@deri.org

1 From Web 1.0 to Web 2.0

The structural and syntactic web put in place in the early 90s is still much the same as what we use today: resources (web pages, files, etc.) connected by untyped hyperlinks. By untyped, we mean that there is no easy way for a computer to figure out what a link between two pages means - for example, on the W3C website, there are hundreds of links to the various organisations that are registered members of the association, but there is nothing explicitly saying that the link is to an organisation that is a "member of" the W3C or what type of organisation is represented by the link. On John's work page, he links to many papers he has written, but it does not explicitly say that he is the author of those papers or that he wrote such-and-such when he was working at a particular university.

In fact, the Web was envisaged to be much more, as one can see from the image in Fig. 1 which is taken from Tim Berners-Lee's original outline for the Web in 1989, entitled "Information Management: A Proposal". In this, all the resources are connected by links describing the type of relationships, e.g. "wrote", "describe", "refers to", etc. This is a precursor to the Semantic Web which we will come back to later.

Web 2.0 is a widely used and wide-ranging term (in terms of interpretations) made popular by Tim O'Reilly². But what exactly is it? If you ask ten different people you'll probably come up with at least five answers. One source³ says that "Web 2.0 ... has ... come to refer to what some people describe as a second phase of architecture and application development for the World Wide Web." You can think of it as a web where "ordinary" users can meet, collaborate, and share [content] using social software applications on the

* This material is based upon works supported by Science Foundation Ireland Grant No. SFI/02/CE1/I131.

² <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

³ http://en.wikipedia.org/wiki/Web_2

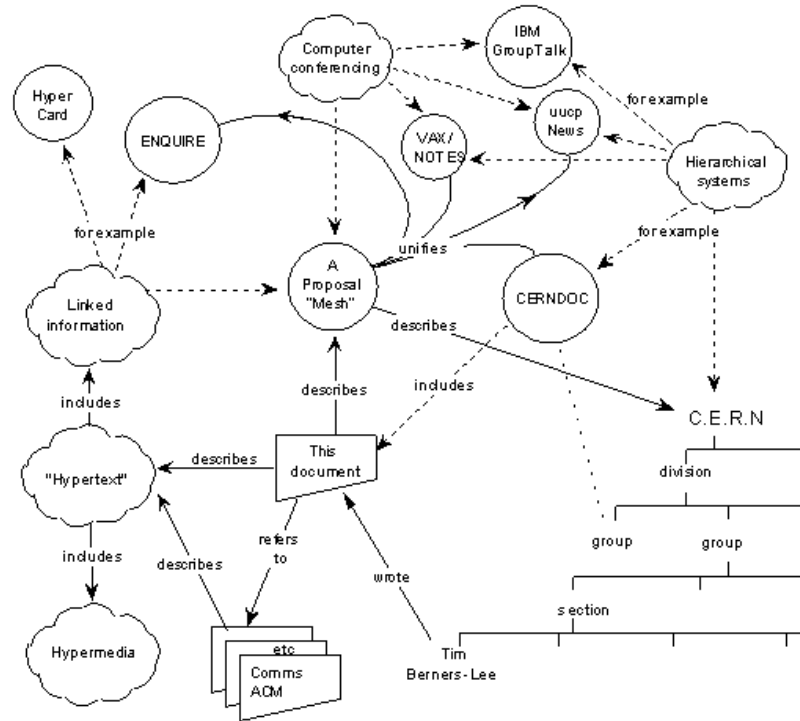


Fig. 1. From Tim Berners-Lee's early proposal for the World Wide Web (<http://www.w3.org/History/1989/proposal.html>).

Web - via tagged items, social bookmarking, AJAX functionality, etc. There are many popular examples that work along this collaboration and sharing meme: Bebo, del.icio.us, digg, Flickr, UseAMap.com, Technorati, orkut, 43 Things, Wikipedia, etc.

Over the last 13 years, there has been a shift from just 'existing' on the Web to participating on the Web. Web 2.0 is a platform for social and collaborative exchange with reusable community contributions, where anyone can mass-publish using web-based social software and others can subscribe to desired information, news, data flows, or other services. It is "social software" that is being used for this communication and collaboration, software that "lets people rendezvous, connect or collaborate by use of a computer network. It results in the creation of shared, interactive spaces..." Examples include instant messaging, IRC, forums, blogs, wikis, SNS (social network services), social bookmarking, podcasts, and MMOGs or MMORPGs.

O'Reilly wrote on the seven features or principles of Web 2.0, to which some have added an eighth: the long tail phenomenon⁴. But in short, Web 2.0

⁴ http://en.wikipedia.org/wiki/Long_tail

is all about being more open, more social, and through user-created content, cheaper.



Fig. 2. The main aspects of Web 2.0.

2 Tags, Folksonomies, Vocabularies and Web 3.0

A key feature of Web 2.0 sites is community-contributed content that may be tagged and can be commented on by others. That content can be virtually anything: blog entries, board posts, videos, audio, images, wiki pages, user profiles, bookmarks, events, etc. Soon we will see sites with live multiplayer video games appearing in little browser-embedded windows just YouTube does videos, with running commentaries going on about the games in parallel. Tagging is common to many Web 2.0 sites - a tag is a keyword that acts like a subject or category for the associated content. Then we have folksonomies: collaboratively generated, open-ended labeling systems that enable Web 2.0 users to categorise content using the tags system, and to thereby visualise popular tag usages via "tag clouds" (visual depictions of the tags used on a particular website, like a weighted list in visual design).

Folksonomies are one step in the same direction as what some have termed Web 3.0⁵, or the Semantic Web. (The Semantic Web often uses topdown controlled vocabularies to describe various domains, but can also utilise folksonomies and therefore develop more quickly since folksonomies are a great big distributed classification system with low entry costs.) As Tim-Berners Lee et al. said in *Scientific American* in 2001, the Semantic Web is "an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation". The word "semantic" stands for "the meaning of", and therefore the Semantic Web is one that is able to describe things in a way that computers can better understand. Some of the more popular Semantic Web vocabularies include FOAF (Friend-of-a-Friend, for social networks), SIOC (for online communities and content), and Geo (for geographic locations).

⁵ <http://turing.cs.washington.edu/NYT-KnowItAll.htm>

As already mentioned in the previous chapter, the Semantic Web consists of metadata that is associated with web resources, and then there are associated vocabularies or "ontologies" that describe what this metadata is and how it is all related to each other. SEO experts have known that adding metadata to their websites can often improve the percentage of relevant document hits in search engine result lists, but it is hard to persuade web authors to add metadata to their pages in a consistent, reliable manner (either due to perceived high entry costs or because it is too time consuming). For example, few web authors make use of the simple Dublin Core metadata system, even though the use of DC meta tags can increase their pages' prominence in search results.

The main power of the Semantic Web lies in interoperability, and combinations of vocabulary terms: interoperability and increased connectivity is possible through a commonality of expression; vocabularies can be combined and used together: e.g. a description of a book using Dublin Core metadata can be augmented with specifics about the book author using the FOAF vocabulary. Vocabularies can also be easily extended (modules, etc.). Through this, true intelligent search with more granularity and relevance is possible: e.g. a search can be personalised to an individual by making use of their identity profile and relationship information.

The challenge for the Semantic Web is related to the chicken-and-egg problem: it is difficult to produce data without interesting applications, and vice versa. The Semantic Web cannot work all by itself, because if it did it would be called the "Magic Web". For example, it is not very likely that you will be able to sell your car just by putting your a Semantic Web file on the Web. Society-scale applications are required, i.e. consumers and processors of Semantic Web data, Semantic Web agents or services, and more advanced collaborative applications that make real use of shared data and annotations.

The Semantic Web effort is mainly towards producing standards and recommendations that will interlink applications, and the primary Web 2.0 meme as already discussed is about providing user applications. These are not mutually exclusive⁶: with a little effort, many Web 2.0 applications can and do use Semantic Web technologies to great benefit, and the picture in Fig. 3 shows some evolving areas where these two streams have and will come together: semantic blogging, semantic wikis, semantic social networks and the Semantic Desktop all fall in the realm of what he terms the Metaweb, or "social semantic information spaces". Semantic MediaWiki⁷, for example, has already been commercially adopted⁸ by Centiare (now MyWikiBiz).

There are also great opportunities for mashing together of both Web 2.0 data or applications and Semantic Web technologies, which just require the use of some imagination. Dermod Moore wrote of one such Web 2.0 applica-

⁶ http://www.oreillynet.com/xml/blog/2005/10/is_web_20_killing_the_semantic.html

⁷ http://meta.wikimedia.org/wiki/Semantic_MediaWiki

⁸ <http://www.sbwire.com/news/view/9912>

| Web 1.0 | Web 2.0 | Web 3.0 |
|---|--------------------------------------|---|
| Personal Websites | Blogs | Semantic Blogs: semiBlog, Haystack, Semblog, Structured Blogging |
| Content Management Systems, Britannica Online | Wikis, Wikipedia | Semantic Wikis: Semantic MediaWiki, SemperWiki, Platypus, dbpedia, Rhizome |
| Altavista, Google | Google Personalised, DumbFind, Hakia | Semantic Search: SWSE, Swoogle, Intellidimension |
| CiteSeer, Project Gutenberg | Google Scholar, Book Search | Semantic Digital Libraries: JeromeDL, BRICKS, Longwell |
| Message Boards | Community Portals | Semantic Forums and Community Portals: SIOC, OpenLink DataSpaces |
| Buddy Lists, Address Books | Online Social Networks | Semantic Social Networks: FOAF, PeopleAggregator |
| ... | ... | Semantic Social Information Spaces: Nepomuk, Gnowsis |

Fig. 3. From Web 1.0 to Web 3.0.

tion mashing for a hobby project⁹: a Scuttle¹⁰ + Gregarius¹¹ + Feedburner¹² + Grazr¹³ hybrid that allows one to aggregate ones favourite blogs or other content on a particular topic and then to annotate bookmarks to the most interesting content found. Bringing this a step further, we could have a "semantic social collaborative resource aggregator. In this hypothetical system:

- Social network members specify their favourite content sources
- You and your friends specify any topics of interest
- You specify friends whose topic lists you value
- Metadata aggregator collects content from sites you and friends like (which may be human tagged, or could be auto-tagged)
- Highlights content that may be of interest to you or your friends
- If nothing of interest is currently available, content sources may have semantically-related sources in other communities for secondary content acquisition and highlighting
- You bookmark and tag the interesting content, and share!

We will now discuss three of the most popular Web 2.0 application areas: blogs, wikis and online social networks, and describe how each of these can be enhanced with semantics to not only provide more functionality but also to create an overall interconnected set of social information spaces.

⁹ <http://bonhom.ie/2006/04/what-weeks-delay-can-produce.html>

¹⁰ <http://scuttle.org/>

¹¹ <http://gregarius.net/>

¹² <http://www.feedburner.com/>

¹³ <http://grazr.com/>

3 Blogging and Semantic Publishing

Blogs are websites which contain periodic time-stamped posts (in reverse chronological order) about a particular genre or touching on a number of topics of interest. They range from individual's online diaries or journals to promotional tools used by companies or political campaigns, and many allow public commenting on their posts. They are also starting to cross the generation gap - teenagers might have a blog via a social networking service, their parents may blog themselves and perhaps grandparents could also be posting, reading or commenting on posts.

The growth and takeup of blogs over the past four years has been dramatic, with a doubling in the size of the blogosphere every six or so months (according to statistics from Technorati¹⁴). Over 120,000 blogs are created every day, working out at about one a second. Nearly 1.5 million blog posts are being made each day, with over half of bloggers contributing to their sites three months after the blog's creation.

Similar to accidentally wandering onto message boards and web-enabled mailing lists, when searching for something on the Web, one often happens across a relevant entry on someone's blog. RSS feeds are also a useful way of accessing information from your favourite blogs, but they are usually limited to the last 15 entries, and do not provide much information on exactly who wrote or commented on a particular post, or what the post is talking about. Some approaches like SIOC¹⁵ aim to enhance the semantic metadata provided about blogs, forums and posts, but there is also a need for more information about what exactly a person is writing about. When searching for particular information in or across blogs, it is often not that easy to get it because of "splogs" (spam blogs) and also because of the fact that the virtue of blogs so far has been their simplicity - apart from the subject field, everything and anything is stored in one big text field for content. Keyword searches may give some relevant results, but useful questions such as "find me all the restaurants that bloggers reviewed in Dublin with a rating of at least 5 out of 10" cannot be posed, and you cannot easily drag-and-drop events or people or anything (apart from URLs) mentioned in blog posts into your own applications.

There have been some approaches to tackle this issue of adding more information to posts, so that queries can be made and the things that people talk about can be reused in other posts or applications (because not everyone is being served well by the lowest common denominator that we currently have in blogs). One approach is called "structured blogging"¹⁶ and the other "semantic blogging".

Structured blogging is an open source community effort that has created tools to provide microcontent (including microformats¹⁷ like hReview) from

¹⁴ <http://technorati.com/weblog/2007/04/328.html>

¹⁵ <http://sioc-project.org/>

¹⁶ <http://structuredblogging.org/>

¹⁷ <http://microformats.org/>

popular blogging platforms such as WordPress and Moveable Type. In structured blogging, packages of structured data are becoming post components. Sometimes (not all of the time) a person will have a need for more structure in their posts - if they know a subject deeply, or if their observations or analyses recur in a similar manner throughout their blog - then they may best be served by filling in a form (which has its own metadata and model) during the post creation process. For example, someone may be writing a review of a film they went to see, or reporting on a sports game they attended, or creating a guide to tourist attractions they saw on their travels. Not only do people get to express themselves more clearly, but blogs can start to interoperate with enterprise applications through the microcontent that is being created in the background.

Take the scenario where someone (or a group of people) is reviewing some soccer games that they watched. Their after-game soccer reports will typically include information on which teams played, where the game was held and when, who were the officials, what were the significant game events (who scored, when and how, or who received penalties and why, etc.) - it would be easier for these blog posters if they could use a tool that would understand this structure, presenting an editing form with the relevant fields, and automatically create both HTML and RSS with this structure embedded in it. Then, others reading these posts could choose to reuse this structure in their own posts, and their blog reader or creator could make this structure available when the blogger is ready to write. As well as this, reader applications could begin to answer questions based on the form fields available - "show me all the matches from Germany with more than two goals scored", etc.

At the moment, structured blogging tools (such as those from LouderVoice¹⁸) provide a fixed set of forms that bloggers can fill in for things like reviews, events, audio, video and people - but there is no reason that people could not create custom structures, and news aggregators or readers could auto-discover an unknown structure, notify a user that a new structure is available, and learn the structure for reuse in the user's future posts.

Semantic Web technologies can also be used to ontologise any available post structures for more linkage and reuse. Blog posts are usually only tagged on the blog itself by the post creator, using free-text keywords such as "scotland", "movies", etc. (unless they are bookmarked and tagged by others using social bookmarking services like del.icio.us¹⁹ or personal aggregators like Gregarius²⁰. Technorati, the blog search engine, aims to use these keywords to build a "tagged web". Both tags and hierarchial categorisations of blog posts can be further enriched using the SKOS²¹ framework. However, there is often much more to say about a blog post than simply what category it belongs in.

¹⁸ <http://www.loudervoice.com/>

¹⁹ <http://del.icio.us/>

²⁰ <http://gregarius.net/>

²¹ <http://en.wikipedia.org/wiki/SKOS>

This is where semantic blogging comes in. Traditional blogging is aimed at what can be called the "eyeball Web" - i.e. text, images or video content that is targetted mainly at people. Semantic blogging aims to enrich traditional blogging with metadata about the structure (what relates to what and how) and the content (what is this post about - a person, event, book, etc.). In this way, metadata-enriched blogging can be better understood by computers as well as people.

In structured blogging, microcontent such as microformats²² is positioned inline in the HTML (and subsequent syndication feeds) and can be rendered via CSS. Structured blogging and semantic blogging do not compete, but rather offer metadata in slightly different ways (using microcontent like microformats and RDF²³ respectively). There are already mechanisms such as GRDDL²⁴ which can be used to move from one to the other.

The question remains as to why one would choose to enhance their blogs and posts with semantics. Current blogging offers poor query possibilities (except for searching by keyword or seeing all posts labelled with a particular tag). There is little or no reuse of data offered (apart from copying URLs or text from posts). Some linking of posts is possible via direct HTML links or trackbacks, but again, nothing can be said about the nature of those links (are you agreeing with someone, linking to an interesting post, or are you quoting someone whose blog post is directly in contradiction with your own opinions?). Semantic blogging aims to tackle some of these issues, by facilitating better (i.e. more precise) querying when compared with keyword matching, by providing more reuse possibilities, and by creating "richer" links between blog posts. It is not simply a matter of adding semantics for the sake of creating extra metadata, but rather a case of being able to reuse what data a person already has in their desktop or web space and making the resulting metadata available to others. People are already (sometimes unknowingly) collecting and creating large amounts of structured data on their computers, but this data is often tied into specific applications and locked within a user's desktop (e.g. contacts in a person's addressbook, events in a calendaring application, author and title information in documents, audio metadata in MP3 files). Semantic blogging can be used to "lift" or release this data onto the Web.

For example, looking at the picture in Fig. 4 from the semiBlog documentation by Knud Möller, Aidan writes a blog post which he annotates using content from his desktop calendaring and addressbook applications. He publishes this post onto the Web, and John, reading this post, can reuse the embedded metadata in his own desktop applications. In this picture, the semantic blog post is being created by annotating a part of the post text about a person with an address book entry that has extra metadata describing that person. Once a blog has semantic metadata, it can be used to perform queries

²² <http://microformats.org/>

²³ <http://www.w3.org/RDF/>

²⁴ <http://www.w3.org/2004/01/rdxh/spec>

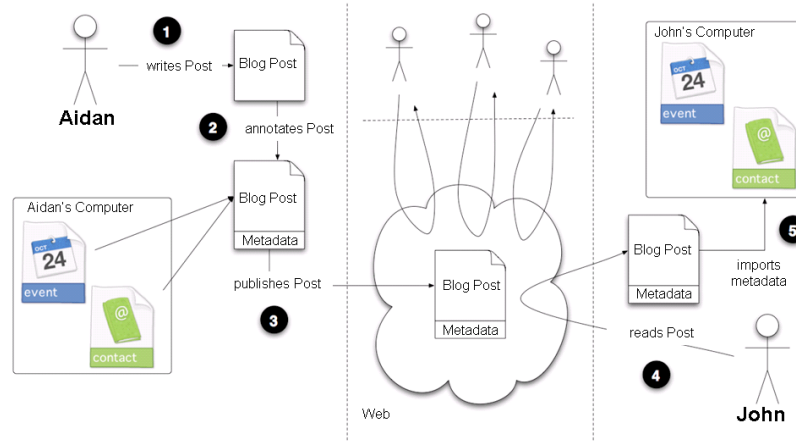


Fig. 4. Lifting semantic data from the desktop to the Web and back again.

such as "which blog posts talk about papers by Stefan Decker?"; it can be used for browsing not only across blogs but also other kinds of discussion methods; or it can be used by blog readers for importing metadata into desktop applications (or using the Web as a clipboard). As well as semiBlog, other semantic blogging systems have been developed by HP²⁵, the National Institute of Informatics, Japan²⁶ and MIT²⁷.

4 Using Wikis for Creating Structured Documents

It is not just blog posts that are being enhanced by structured metadata and semantics - this is happening in many other Web 2.0 application areas. Wikis such as the Wikipedia have contained structured metadata in the form of templates for some time now, and at least twenty "semantic wikis" have also appeared to address a growing need for more structure in wikis.

Many people are familiar with the Wikipedia²⁸, but less know exactly what a wiki is. In short, a wiki is an "information space" (web or desktop application) that allows users to easily add and edit content, and is especially suited for collaborative writing. Wikis rely on cooperation, on checks and balances of the wiki site members, and a belief in the sharing of ideas. The name comes from a Hawaiian phrase, "wiki wiki", which means to hasten or go quickly. Ward Cunningham created the first wiki in 1995, and wikis are now being used for free dictionaries, book repositories, event organisation, and

²⁵ http://www.hpl.hp.com/personal/Steve_Cayzer/semblog.htm

²⁶ <http://www.semblog.org/>

²⁷ <http://theory.csail.mit.edu/dquan/iswc2004-blog.ppt>

²⁸ <http://www.wikipedia.org/>

software development. They have become increasingly used in enterprise environments for collaborative purposes: research projects, papers and proposals, coordinating meetings, etc. Ross Mayfield's SocialText²⁹ produced the first commercial open source wiki solution, and many companies now use wikis as one of their main intranet collaboration tools. There are a plethora (hundreds) of wiki software systems now available, ranging from MediaWiki³⁰, the software used on the Wikimedia family of sites, and Eugene Eric Kim's PurpleWiki³¹, where fine grained elements on a wiki page are referenced by purple numbers, to Alex Schröder's OddMuse³², a single Perl script wiki install, and WikidPad³³, a desktop-based wiki for managing personal information. Many are open source, free, and will often run on multiple operating systems. The differences between wikis are usually quite small but can include the development language used (Java, PHP, Python, Perl, Ruby, etc.), the database required (MySQL, flat files, etc.), whether attachment file uploading is allowed or not, spam prevention mechanisms, page access controls, RSS feeds, etc.

The Wikipedia project consists of 250 different wikis, corresponding to a variety of languages. The English language one is currently the biggest, with over 2 million pages, but there are wikis in languages ranging from Irish to Arabic to Chinese (and even in constructed languages such as Esperanto). A typical wiki page will have two buttons of interest: "Edit" and "History". Normally, anyone can edit an existing wiki article, and if the article does not exist on a particular topic, anyone can create it. If someone messes up an article (either deliberately or erroneously), there is a revision history so that the contents can be reverted or fixed. There is a certain amount of ego-related motivation in contributing to a wiki - people like to show that they know things, to fix mistakes and fill in gaps in underdeveloped articles (stubs), and to have a permanent record of what they have contributed via their registered account. By providing a template structure to input facts about certain things (towns, people, etc.), wikis also facilitate this user drive to populate wikis with information.

For some time on the Wikipedia and in other wikis, templates have been used to provide a consistent look to the content placed within article texts. They can also be used to provide a structure for entering data, so that it is easy to extract metadata about the topic of an article (e.g. from a template field called "population" in an article about London). Semantic wikis³⁴ bring this to the next level by allowing users to create semantic annotations anywhere within a wiki article text for the purposes of structured access and finer-grained searches, inline querying, and external information reuse. There are

²⁹ <http://www.socialtext.com/>

³⁰ <http://www.mediawiki.org>

³¹ <http://www.blueoxen.org/tools/purplewiki/>

³² <http://www.oddmuse.org/>

³³ <http://www.jhorman.org/wikidPad/>

³⁴ http://en.wikipedia.org/wiki/Semantic_wiki

already about 20 semantic wikis in existence, and one of the largest ones is Semantic MediaWiki³⁵, based on the popular MediaWiki system.

Let us take an example of providing structured access to information in wikis. There is a Wikipedia page about JK Rowling that has a link to "Harry Potter and the Philosopher's Stone" (and to other books that she has written), to Edinburgh because she lives there, and to Scholastic Press, her publisher. In a traditional wiki, you cannot perform fine-grained searches on the Wikipedia dataset such as "show me all the books written by JK Rowling", or "show me all authors that live in the UK", or "what authors are signed to Scholastic", because the type of links (i.e. the relationship type) between wiki pages are not defined. In Semantic MediaWiki, you can do this by linking with `[[author of::Harry Potter and the Philosopher's Stone]]` rather than just the name of the novel. There may also be some attribute such as `[[birthdate:=1965-07-31]]` which is defined in the JK Rowling article. Such attributes could be used for answering questions like "show me authors over 40" or for sorting articles.

Some semantic wikis also provide what is called inline querying. A question such as `"?page dc:creator EyalOren"` (or find me all pages where the creator is Eyal Oren) is processed as a query when the page is viewed and the results are shown in the wiki page itself. Also, when defining some relationships and attributes for a particular article (e.g. `"foaf:gender Male"`), other articles with matching properties can be displayed along with the article. Finally, just as in the semantic blogging scenario, wikis can enable the Web to be used as a clipboard, by allowing readers to drag structured information from wiki pages into other applications (for example, geographic allowing readers to drag structured information from wiki pages into other applications (for example, geographic data about locations on a wiki page could be used to annotate information on an event or a person in your calendar application or address book software respectively).

5 Social Networking Services and their Features

Social networking services (SNS) allow a user to create and maintain an online network of close friends or business associates for social and professional reasons. There has been an explosion in the number of online social networking services in the past five years, so much so that the terms YASN and YASNS (Yet Another Social Network[ing Service]) have become commonplace. But these sites do not usually work together and therefore require you to re-enter your profile and redefine your connections when you register for each new site.

The "friend-of-a-friend effect" often occurs when someone tells someone something and they then tell you - linked to the theory that anybody is connected to everybody else (on average) by no more than six degrees of separation. This number of six degrees came from a sociologist called Stanley

³⁵ <http://semantic-mediawiki.org/>

Milgram who conducted an experiment in the late 1960s. Random people from Nebraska and Kansas were told to send a letter (via intermediaries) to a stock broker in Boston. However, they could only give the letter to someone that they knew on a first-name basis. Amongst the letters that found their target (around 20 percent), the average number of links was around 5.5 (rounded up to 6). Some argue that this figure should be much lower due to the low percentage of letters that actually arrived. Some other related ideas include the Erdős number (the number of links required to connect scholars to mathematician Paul Erds³⁶, a prolific writer who co-authored over 1500 papers with more than 500 authors), and the Kevin Bacon game (the goal is to connect any actor to Kevin Bacon³⁷, by linking actors who have acted in the same movie). The six degrees idea is nicely summed up by this quote from a film called "Six Degrees of Separation" written by John Guare:

"I read somewhere that everybody on this planet is separated by only six other people. Six degrees of separation between us and everyone else on this planet. The President of the United States, a gondolier in Venice, just fill in the names. [...] It's not just big names – it's anyone. A native in a rain forest, a Tierra del Fuegan, an Eskimo. I am bound – you are bound – to everyone on this planet by a trail of six people."

It is often found that even though one route is followed to get in contact with a particular person, after talking to them there is another obvious connection that was not previously known about. This is part of the small-world network³⁸ theory, which says that most nodes in a network exhibiting small-world characteristics (such as a social network) can be reached from every other node by a small number of hops or steps.

There are now many websites acting as a social networking service. The idea behind such services is to make people's real-world relationships explicitly defined online - whether they be close friends, professional colleagues or just people with common interests. Most SNSs allow one to surf from a list of friends to find friends-of-friends, or friends-of-friends-of-friends for various purposes. SNSs have become the new digital public places of Web 2.0 - there has been a huge takeup of sites such as MySpace, LinkedIn, Bebo and Facebook. Most SNSs allow content generation and sharing, and there is also a gradual transformation of SNSs into public e-markets - either through product promotions or targetted ads.

Social networking services usually offer the same basic functionalities: network of friends listings (showing a person's "inner circle"), person surfing, private messaging, discussion forums or communities, events management, blogging, commenting (sometimes as endorsements on people's profiles), and

³⁶ <http://www.oakland.edu/enp/>

³⁷ <http://oracleofbacon.org/>

³⁸ http://en.wikipedia.org/wiki/Small-world_network

media uploading. Some motivations for SNSs include building friendships and relationships, arranging offline meetings, curiosity about others, arranging business opportunities, or job hunting. People may want to meet with local professionals, create a network for parents, network for social (dating) purposes, get in touch with a venture capitalist, or find out if they can link to any famous people via their friends.

Before 2002, most people networked using services such as OneList, ICQ or eVite. The first big SNS in 2002 was Friendster; in 2003, LinkedIn (a SNS for professionals) and MySpace (target audience is 20-30 years) appeared; then in 2004, orkut (Google's SNS) and Facebook (by a college student for college students) were founded; these were followed by Bebo (target audience is 10-20 years) in 2005. There has also been a lot of venture capital investment in and sales of social networking services as well. An interesting statistic related to this is that as of November 2006, the ten most popular domains accounted for about 40 percent of all page views on the Web, and nearly half of these page views were from SNSs (MySpace and Facebook).

Even in a small-sized SNS, there can be a lot of links available for analysis, and this data is usually meaningless when viewed as a whole, so one needs to apply some social network analysis (SNA) techniques. Apart from textbooks³⁹, there are many academic resources for social networks⁴⁰ and SNA⁴¹. For example, the tool Pajek⁴² can be used to drill down into various networks. A common method is to reduce the amount of relevant social network data by clustering. One could choose to cluster people by common friends, by shared interests, by geography, by tags, etc. In social network analysis, people are modelled as nodes or "actors". Relationships (such as acquaintanceship, co-authorship, friendship, etc.) between actors are represented by lines or edges. This model allows analysis using existing tools from mathematical graph theory and mapping, with target domains such as movie actors, scientists and mathematicians (as already mentioned), sexual interaction, phone call patterns or terrorist activity. There are some nice tools for visualizing these models, such as Vizster⁴³ by Heer and Boyd, based on the Prefuse⁴⁴ open-source toolkit. Others have combined SNA with Semantic Web technologies⁴⁵ to determine social behaviour patterns, and MIT Media Lab are conducting mobile SNA research via their "reality mining⁴⁶" project. On the security front, the NSA are using social network analysis technologies for homeland

³⁹ <http://www.amazon.com/Social-Network-Analysis-Applications-Structural/dp/0521387078>

⁴⁰ <http://www.socialnetworks.org/>

⁴¹ <http://lrs.ed.uiuc.edu/tseportal/analysis/social-network-analysis/>

⁴² <http://vlado.fmf.unilj.si/pub/networks/pajek/>

⁴³ <http://jheer.org/vizster/>

⁴⁴ <http://prefuse.org/>

⁴⁵ <http://www.blogninja.com/galway-iswc2005.ppt>

⁴⁶ <http://reality.media.mit.edu/>

security, and there have been reports from the New Scientist of "automated intelligence profiling"⁴⁷ from sites like MySpace.

So what does the future hold for SNS sites? It has been theorised⁴⁸ that many sites only work where there is some "object-centered sociality" in networks, i.e. users are connected via a common object, e.g. their job, university, hobby, etc. In this way, it is probable that people's SNS methods will move closer towards simulating their real-life social interaction, so that people will meet others through something they have in common, not by randomly approaching each other. In the future, we will no doubt see better interaction methods with friends a'la Second Life.

Another future requirement is for distributed social networks and reusable profiles. There have been a lot of complaints about the walled gardens that are social network sites. Some of the most popular SNSs would not exist without the walled garden approach, but some flexibility would be useful. Users may have many identities on different social networks, where each identity was created from scratch. A reusable profile would allow a user to import their existing identity and connections (from their own homepage or from another site they are registered on), thereby forming a single global identity with different views (e.g. there is Videntity⁴⁹ which works with OpenID⁵⁰ and FOAF⁵¹).

⁴⁷ <http://www.newscientist.com/article/mg19025556.200?DCMP=NLC-letter&nsref=samplemg19025556.200>

⁴⁸ http://www.zengstrom.com/blog/2005/04/why_some_social.html

⁴⁹ <http://videntity.org/>

⁵⁰ <http://openid.net/>

⁵¹ <http://foaf-project.org/>