

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA

MECANISMO DE SUGESTÃO E
PROCESSO DE CARACTERIZAÇÃO
DE REDES SOCIAIS

TRABALHO DE GRADUAÇÃO DO CURSO DE CIÊNCIA DA COMPUTAÇÃO

Aluno – Nitai Bezerra da Silva, nbs@cin.ufpe.br

Orientador - Tsang Ing Ren, tir@cin.ufpe.br

Recife, 23 de janeiro de 2008.

MECANISMO DE SUGESTÃO E PROCESSO DE CARACTERIZAÇÃO DE REDES SOCIAIS

Trabalho de conclusão de curso apresentado
como parte das atividades para obtenção do
título de graduado, do curso de ciência da
computação da Universidade Federal de
Pernambuco

Prof^o orientador: Tsang Ing Ren

Recife, 2008

AGRADECIMENTOS

*Agradeço a minha filha por ter
chegado nessa fase tão conturbada trazendo
alegria e serenidade ao meu coração. Também
agradeço a Davi Pires Rios pelo empenho na
concepção do projeto que resultou no meu
objetivo. Agradeço a Sonia Maria Bezerra
Silva, minha mãe, por fornecer o balanceado
almoço de cada dia.*

RESUMO

Nos últimos anos o crescente público em sistemas de redes sociais mudou o rumo das principais empresas da área de informática. Oferecendo maior interatividade e conteúdo de fácil acesso, eles são resultado da junção e convergência de vários outros sistemas menores de propósitos específicos, como por exemplo: álbuns de fotos, blogs, e-mails, mensageiros eletrônicos (MSN, ICQ, Gtalk), além de outros. Com tantos serviços e qualidade, as redes sociais têm sido mais utilizadas do que os sistemas de e-mail. Como o próprio nome já diz, as redes sociais promovem a socialização das pessoas. Tem como maior conteúdo as estruturas que representam pessoas e os dados individuais dos perfis. Como maior serviço, a ferramenta para mapear o círculo de amigos e navegar através dele.

Utilizando modelos de redes complexas como: *small world* e *scale free*, esse projeto propõe um mecanismo para ajudar cada usuário a aumentar seu círculo de amigos. A partir de um mecanismo de sugestão inteligente, os usuários encontrarão com maior rapidez as pessoas que devem incluir em seu círculo de amigos. Essa ferramenta de busca faz uso da estrutura social já existente que foi criada pelas pessoas mais próximas do usuário, e utilizando algumas métricas podem-se avaliar certas forças de relacionamento. O bom funcionamento dela depende da qualidade da estrutura social das pessoas mais próximas, classificando ela como ferramenta colaborativa.

Palavras-chave: sugestão, rede social, redes complexas, small world, scale free, busca, colaboração.

ABSTRACT

In the last years we have witness of people interest in social network systems. This fact has changed the trend of research interest of the main companies related. Providing more interactive and easy content access, social network systems are a result from the merging and convergence of several others smaller systems with specific purpose, i.e. album of photo, blogs, e-mails, instant messenger (MSN, ICQ, Gtalk), and others. With so much services and quality, social network has been used more often than the e-mail systems. Like the name says, the social networks promotes a way to get acquainted to each other. They have the network structure, that means, peoples and the individual profile data as the best content. And as the best service tools for mapping friends and navigate through relationships.

Using complex network models like small world and scale free, this project proposes a mechanism to help each user to increase they friend's list. From a smart mechanism to suggest, the users will find faster the other users to add in their friend's list. This search tool uses the existing social structure that was created by the closest peoples of the user, and metrics for some relationship power interaction that can be evaluated. Its proper function depends on the social network's quality from the closest people, classifying it as collaborative tool.

Key-words: suggestion, social network, complex systems, small world, scale free, search, collaboration.

Sumário

1. Redes.....	8
1.1. Redes Sociais.....	8
1.2. Redes Complexas	11
1.2.1. Sistemas Complexos	11
1.2.2. Estudo de Redes	12
1.3. Busca em redes complexas	18
1.3.1. Aplicabilidade	21
2. Desenvolvimento	22
2.1. Construindo a rede social	23
2.2. Caracterização	25
2.2.1. Caracterizando a rede como <i>scale free</i>	26
2.2.2. Caracterizando a rede como <i>smal -world</i>	28
2.3. Mecanismo de sugestão	35
2.3.1. Filtrar.....	35
2.3.2. Ordenar.....	36
3. Conclusões e trabalhos futuros	44
Bibliografia.....	46

Introdução

Assim como aconteceu a bolha da informática no final dos anos 90, atualmente vivemos numa nova fase de inchamento do mercado. Este inchamento está sendo alavancado por sistemas Web que oferecem serviços de interatividade. Dentre eles, podemos destacar as redes sociais como serviço mais bem sucedido. Algumas redes sociais possuem mais de 100.000.000 de usuários e têm se tornado portais que agregam uma variedade de serviços de comunicação e troca de conteúdo. A qualidade da rede social, assim como a capacidade de interatividade está diretamente relacionada à quantidade de relacionamentos por usuário. Mais precisamente, está relacionada com a equivalência do círculo social de um usuário com o mesmo círculo no mundo real. Se todos estão na rede social, então é dispensável a utilização de outros vários meios de comunicação.

O objetivo deste trabalho é apresentar um mecanismo de sugestão de relacionamentos e seus benefícios singulares e colaborativos. São utilizados modelos de redes complexas para caracterizar e qualificar as soluções implementadas. Para o desenvolvimento experimental, foi criado um framework constituído de um módulo de extração baseado em crawlers, uma base de dados robusta para suportar pesados procedimentos, um módulo de scripts de testes, além do módulo de sugestão.

1. REDES

1.1. Redes Sociais

Uma rede social é uma estrutura social composta por nós (que são geralmente indivíduos ou organizações) que estão ligadas por um ou mais tipos específicos de interdependência, como valores, visões, idéias, trocas financeiras, amigos, parentesco, antipatia, conflito, comércio, links da web, relações sexuais, dispersão de doenças (epidemiologia), rotas aéreas, e outros. [2]

Análises em redes sociais vêem as relações sociais em termos de nós e ligações. Os nós são os atores individuais dentro da rede, e as ligações são as relações entre os atores. Podem existir muitos tipos de ligações entre os nós. Pesquisadores em várias áreas acadêmicas têm mostrado que redes sociais operam em muitos níveis, desde famílias subindo até o nível de nação. Elas podem representar um papel crítico em determinar o caminho em que problemas são resolvidos, em como organizações funcionam, e em que grau os indivíduos tem sucesso em alcançar seus objetivos.

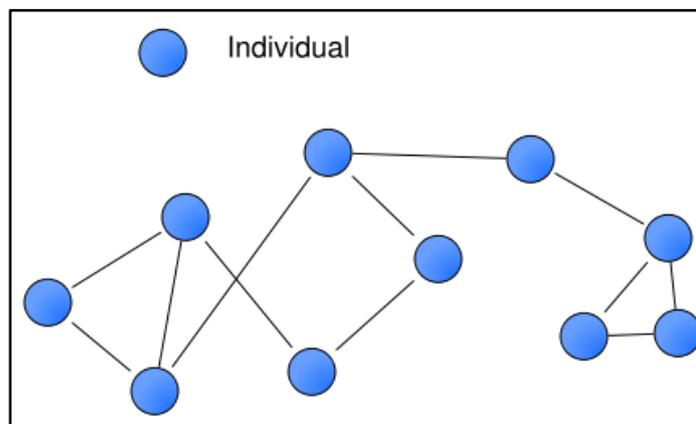


Figura 1 - Exemplo de diagrama de uma rede social

Serviços para redes sociais estão focados na construção e manutenção de redes sociais on-line para comunidades de pessoas que compartilham interesses e atividades,

ou quem estão interessados em explorar os interesses e atividades dos outros, e que necessitam o uso de um software. A maioria dos serviços de redes sociais é baseada na internet e provém um conjunto de várias formas dos usuários interagirem, como bate-papo, mensagens, vídeo, compartilhamento de arquivos, blog, grupos de discussão, etc.. [3]

A noção de que vários computadores ligados eletronicamente poderiam formar as bases para interações sociais mediadas por computador foi sugerido anteriormente por exemplo *The Network Nation* por S. Roxanne Hiltz and Murray Turoff [4] (Addison-Wesley, 1978, 1993) efetivamente esboçou como comunicação mediada por computador – como a internet – deveriam ser desenvolvidas para esse propósito.

Com um pouco mais de uma década de existência, sítios com propósitos de redes sociais incluem Classmates.com (1995), focando as relações entre amigos de turmas de escolas, e SixDegrees.com (1997), focando em ligações indiretas. Dois diferentes modelos de redes sociais que surgiram a partir de 1999 eram baseados em confiança, desenvolvidos pela empresa Epinions.com, e baseados em amizade, como os desenvolvidos por Jonathan Bishop e usado em algumas páginas regionais da Inglaterra entre 1999 e 2001. No Brasil o sistema mais usado até hoje em dia é o Orkut, atualmente mantido pela Google. Porém outros sites com maior sucesso nos EUA como o MySpace.com reportaram ter mais acesso do que as páginas de busca do Google, e o Facebook.com, mais jovem e muito mais promissor tornando-se o sistema web com maior crescimento em quantidade de usuários e valor de mercado.

Em geral, serviços de redes sociais, como MySpace, Facebook, Bebo e Orkut, permitem aos usuários criarem um perfil para si. Usuários podem remeter fotos próprias e podem, freqüentemente, serem amigos de outros usuários. Na maioria dos casos, ambos os usuários precisam confirmar a relação de amizade antes de estarem efetivamente ligados. Por exemplo, se Alice lista Bob como amigo, então Bob deve aprovar a requisição de amizade de Alice antes que eles estejam listados como amigo um do outro. Algumas redes sociais têm características adicionais, como a habilidade de criar grupos que compartilham interesses comuns, remeter vídeos, e manter discussão em fóruns.

O poder da análise de redes sociais se baseia das diferenças dos estudos sociais científicos tradicionais, que assume que os atributos dos atores individualmente – quando eles são amigáveis ou não, inteligentes ou não, etc. – que importa. Esse tipo de

análise produz uma visão diferente, onde os atributos individuais são menos importante do que as relações e ligações com os outros atores dentro da rede. Essa perspectiva tem se mostrado útil para explicar muitos fenômenos reais do mundo. [2]

Redes sociais têm também sido usadas para examinar como organizações interagem umas com as outras, caracterizando as muitas relações informais que ligam os executivos, assim como associações e conexões entre empregados individualmente de diferentes organizações. Por exemplo, poder dentro de uma organização freqüentemente vêm mais do nível em que uma pessoa tem como centro das relações do que o cargo no emprego. Redes sociais também têm um papel chave em contratação, em sucesso comercial, e em desempenho no trabalho.

Abordagens de ARS (Análises em Redes Sociais) e de modelagens de redes tem sido usadas em epidemiologia para ajudar a entender quais os padrões de contato humano ajudam a inibir a difusão de doenças como o HIV numa população. A evolução de redes sociais pode, algumas vezes, ser simulados pelo uso de modelos baseados em agentes, provendo visibilidade dentro da reciprocidade entre regras de comunicação, boatos de doenças e estruturas sociais. Teorias de difusão de inovação exploram redes sociais e seus papéis em influenciar novas idéias e práticas. Mudar os agentes e a opinião dos líderes freqüentemente tem papel mais importante em estimular a adoção de inovação, embora fatores inerentes à inovação também tenham algum papel na influência.

Algumas métricas em análises de redes sociais

“Betweenness”	Grau em que um nó se posiciona entre outros individuais dentro da rede. O perímetro em que um nó está diretamente conectado somente a aqueles outros nós que não estão diretamente conectados uns com os outros; sendo um intermediário; uma ponte.
”Closeness”	O grau que um nó está próximo de todos os outros nós da rede (direta ou indiretamente).
(”Degree”) ”centrality”	O contagem das ligações para outros nós na rede.
”Centralization”	A diferença entre o n de ligações de cada nó dividido pela soma

	máxima possível de diferenças. Uma rede centralizada vai ter muita das suas ligações dispersadas em volta de um ou poucos nós, enquanto uma rede descentralizada é uma em que existe pequena variação entre o n das ligações que cada nó possui.
”Clustering Coefficient”	Uma medida de probabilidade de que duas associações de um nó estão associadas entre si. Um grande valor nessa métrica indica um alto nível de “facções”.
”Path Length”	A distância entre par de nós dentro da rede. A média dessa métrica é a média das mínimas distâncias entre cada par de nós da rede.
”Structural equivalence ”	Refere-se ao perímetro em que nós têm um conjunto comum de nós ligados com outros nós da rede.

Tabela 1 – Métricas utilizadas em redes complexas.

1.2. Redes Complexas

1.2.1. Sistemas Complexos

Exemplos freqüentemente citados de sistemas complexos na natureza e na sociedade incluem o cérebro, o sistema imunológico, células biológicas, redes metabólicas, colônias de formigas, a internet, mercados econômicos, e redes sociais humanas. [4]

Não existe definição formal geralmente aceita de “sistema complexo”. Um sistema complexo pode ser visto como uma grande rede de componentes relativamente simples com nenhum controle central, em que são exibidos comportamentos complexos emergentes. Os termos nessa definição não são rigorosamente definidos. “Componentes relativamente simples” significam que os componentes individuais, ou ao menos suas regras de funcionamento no comportamento coletivo do sistema, são simples com respeito a aqueles comportamentos coletivos. Por exemplo, um único neurônio ou uma única formiga são entidades complicadas na sua própria definição. Embora, o papel da funcionalidade dessas únicas entidades no contexto de um cérebro completo ou numa colônia inteira é relativamente simples se comparada com o comportamento do sistema inteiro.

“Comportamentos complexos emergentes” são mais difíceis de definir. A noção de emergentes refere-se ao fato de que o comportamento global do sistema não é somente complexo, porém surgem das ações coletivas de componentes simples, e o mapeamento de ações individuais para comportamentos coletivos é não-trivial. A noção de não-linearidade é importante aqui: o todo é mais que a soma das partes. A complexidade do comportamento global do sistema está tipicamente caracterizada em termos dos padrões que eles formam, o processamento de informações são adaptativas para o sistema, isto é, aumenta seu sucesso em algum contexto evolucionário ou competitivo. Para caracterizar comportamentos, cientistas de sistemas complexos usam ferramentas de uma variedade de disciplinas, incluindo dinâmica não-linear, teoria da informação, teoria da computação, psicologia comportamental, e biologia evolucionária, entre outras.

Sistemas complexos como implicam seu nome, são tipicamente difíceis de entender. Tradicionalmente uma ciência mais matematicamente orientada como é a física, química, e biologia matemática tem se concentrado em modelos de sistemas simples que são mais entendíveis através da matemática. O surgimento de interesse em entender propriedades gerais dos sistemas complexos tem acompanhado paralelamente o surgimento da computação, por causa do computador tem possibilitado pela primeira vez na história fazer mais modelos aproximados de sistema complexos na natureza.

1.2.2. Estudo de Redes

Recentemente tem existido um forte crescimento no estudo de redes em relação a várias disciplinas, abrangendo desde ciência da computação e comunicação até sociologia e epidemiologia. O estudo matemático de redes surgiu das teorias de grafos, que iniciou no começo do século dezoito com a solução de Euler para o famoso problema “As pontes de Königsberg”. Na década de 50 Erdős e Rényi produziram trabalhos influentes sobre grafos randômicos [5]. Até recentemente, teoria matemática dos grafos não teve um grande impacto nos estudos das redes do mundo real, uma vez que essas propriedades eram totalmente diferentes daquelas dos grafos randômicos. Sociólogos quantitativos têm estudado redes sociais por um longo tempo. Nos últimos 10 anos um grande número de matemáticos aplicados e físicos tem colaborado para desenvolver teorias gerais de redes.

Uma rede (ou grafo) é simplesmente um conjunto de nós (vértices) e ligações (arestas) entre nós. Muitos, talvez a maioria, dos fenômenos naturais podem ser usualmente descrito em termos de rede. O cérebro é uma enorme rede de neurônios ligada pelas sinapses. O controle sobre atividades genéticas dentro de uma celular é uma consequência de uma rede complexa de genes ligados por proteínas regulatórias. Comunidades sociais são redes onde os nós são pessoas (ou organizações de pessoas) dentre tais existem muitos diferentes tipos de relações possíveis.

Aqui estão algumas questões que os cientistas dessa área de redes complexas estão tentando responder:

- Quais medidas topológicas podem ser usadas para caracterizar as propriedades de uma rede?
- Quais propriedades diferentes conjuntos de redes do mundo-real compartilham, e por quê? Como essas propriedades foram descobertas?
- Como nós escrevemos algoritmos eficientes para determinar essas propriedades?
- Como essas propriedades afetam a dinâmica da informação (ou doença, ou outra comunicação) disseminando em tal rede, e a resistência de tal rede a ruídos, falhas de componentes, ou ataques planejados?
- Dado uma rede com certas propriedades, quais são os melhores caminhos para achar nós particulares nessa rede?

Respondendo essas perguntas pode acarretar num grande impacto não somente em nosso entendimento de muitos sistemas naturais e sociais, principalmente na nossa habilidade de criar e usar efetivamente redes complexas, abrangendo desde melhores buscadores na Web e roteadores de internet até o controle da disseminação de doenças, a efetividade de uma organização criminosa, e os resultados e danos ecológicos das ações humanas.

Watts e Barabási descrevem três modelos gerais de redes: *randômicas*, *small-world*, e *scale free* [6]. Esses modelos são cada caracterizado pela forma em que são criados, e em vários resultados estatísticos, como o índice de distribuição, tamanho do caminho médio entre pares e nós, e o índice de agrupamento.

O grau de um nó em uma rede é definido como a quantidade de vizinhos. A distribuição do grau de uma rede e a distribuição da frequência dos diferentes graus dos

nós na rede. Como exemplo, na Figura 2, o nó rotulado n_1 tem grau três, e a distribuição do grau da rede está demonstrado ao lado direito. O índice médio de todos os oito nós é 2.

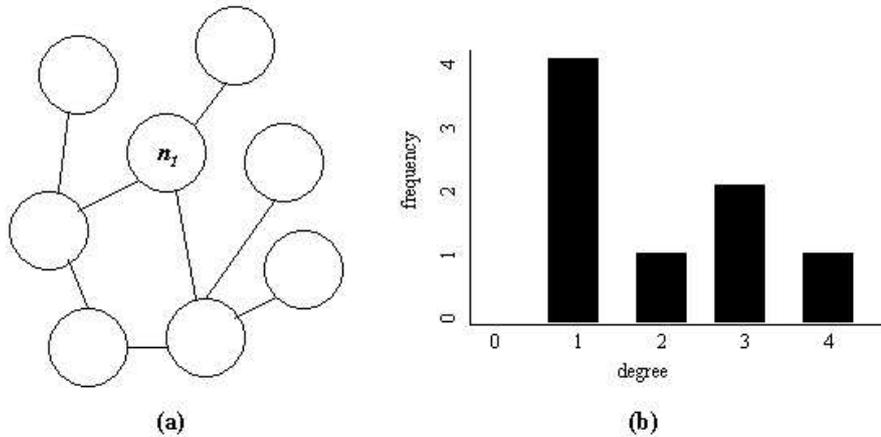


Fig. 2 (a) exemplo de uma rede. (b) distribuição do grau dessa rede.

Uma “rede randômica” é criada especificando que cada par de nós está conectado por uma ligação com probabilidade p uniforme. Tais redes foram estudadas numa visão puramente matemáticas por Erosos e Rényi [5]. No limite para valores grandes de n , muitas das propriedades medianas das redes randômicas podem ser expressas analiticamente, fazendo delas objetos matematicamente elegantes. Algumas propriedades chaves das redes randômicas de Erosos-Rényi mostraram ser completamente diferente daquelas vistas na maioria das redes do mundo-real.

Primeiro, redes randômicas não exibem o forte agrupamento de nós (isto é, sub-redes de nós densamente interconectados) visto no mundo-real. Agrupamento ocorre quando, diz-se, é mais provável você ser um amigo dos amigos dos seus amigos do que de qualquer outra pessoa randomicamente. O índice de agrupamento (*Clustering Coefficient*) de uma rede é a probabilidade média de que dois nós vizinhos de outro dado nó são também vizinhos entre si. O índice de agrupamento de uma rede randômica é exatamente p . De qualquer forma, estudos empíricos das redes do mundo-real freqüentemente mostram índices de agrupamento que são ordens de magnitude maior que os valores sobre os grafos randômicos com o mesmo número de nós e ligações.

Segundo, pode ser demonstrado que o grau de distribuição de uma rede randômica é aproximadamente Gaussiano (na verdade, Poisson, no limite para grandes valores de n), que é muito diferente do grau de distribuição seguindo uma lei de potência (*Power-law*) freqüentemente observada nas redes do mundo-real. Um grau de distribuição

seguinto uma lei de potência $P(k) \sim k^{-\gamma}$, onde $P(k)$ é a probabilidade de um nó ter um grau k , e γ é um número real maior que zero. A Figura 3A mostra uma distribuição de grau seguindo uma lei de potencia com $\gamma = 2.1$ (o grau de distribuição empírico da WWW (World Wide Web)). A distribuição em lei de potência está inclinada para a direita: ele tem uma longa calda para a direita com nenhum ponto de corte (isto é, quando $P(k)$ alcança zero). Em contraste, Figura 3B mostra uma distribuição Gaussiana com uma cauda curto e pontos de corte.

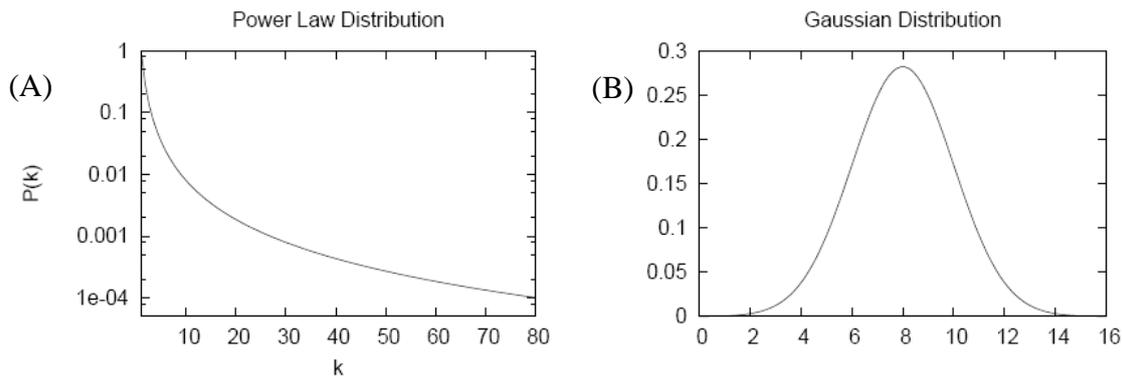


Fig. 3 esquerda: Uma distribuição seguindo uma lei de potencia com $P(k) = k^{-2.1}$, desenhada com escala semi-log, onde k é o grau de um nó e $P(k)$ é a probabilidade de um dado nó tenha aquele grau. **Direita:** Um distribuição Gaussiana hipotética.

1.2.3. Redes small world

Em 1998 Watts e Strogatz introduziram seu modelo de rede *pequeno-mundo* (*small-world*) [7]. Para construir uma rede *small-world* comece com um “anel” do tipo autômato celular de lattice em que cada nó liga-se com k dos seus vizinhos mais próximos. Para cada ligação, com pequena probabilidade p religue uma das extremidades da ligação para qualquer outro nó randomicamente escolhido na rede. Para $p = 0$ o resultado é uma rede completamente regular; para $p = 1$ o resultado é uma rede completamente randômica. Para pequenos, porém não nulos p o resultado é uma rede com muitas conexões locais e pequenas conexões de “longa distâncias”. Tal rede tem as tão conhecidas “propriedades de mundos pequenos”: mesmo existindo algumas conexões de “longas distâncias”, o menor caminho entre dois nós cresce logaritmicamente ou diminui com redes com tamanho n para grau médio fixo. Isso significa que mesmo em redes *small-world* com muitos nós, o menor caminho entre dois nós é relativamente pequeno, explicando a nomenclatura “*small-world*”. Em contraste, numa rede regularmente conectada, distância de caminho cresce linearmente

com o tamanho da rede. Watts e Strogatz descobriram que as propriedades das redes “*small-world*” surgem mesmo quando p , a probabilidade de religar, é muito pequena.

Redes “*small-world*” com pequeno e não nulo valor de p também exibe um alto coeficiente de agrupamento. Em contraste, redes randômicas têm valor médio dos menores caminhos pequeno e coeficiente de agrupamento pequeno. Por outro lado, o modelo de Watts e Strogatz tipicamente resulta em distribuição de grau que não se assemelha com a maioria das redes do mundo-real estudadas.

1.2.4. *Redes scale free*

Na mesma época em que Watts e Strogatz desenvolveram seu modelo de rede *small-world*, Barabási e Albert desenvolveram um modelo de rede alternativo, em que o crescimento via “ligamento preferencial” produzindo as tão conhecidas redes *scale free* [8]. Uma rede *scale free* é simplesmente uma rede com uma *distribuição de grau* seguindo uma *lei de potencia*. Como foi mostrada na parte esquerda da Figura 3. Note que a distribuição Gaussiana mostrada no lado direito da Figura 3 tem valores de corte aonde a distribuição vai para zero. Isso dá à distribuição uma escala intrínseca, por exemplo, se fosse desenhado um gráfico de altura de humano adulto pela sua frequência na população, nenhum humano seria menor que 30 cm ou maior que 4,5 m. Em contraste, uma distribuição *scale free* não tem ponto de corte; instancias de todas as escalas estão presentes, denominando *scale free* (livre de escala).

Numa rede com uma distribuição de grau Gaussiana, a vasta maioria de nós tem aproximadamente o mesmo numero de ligações. É muito raro achar grande desvio da média. Em contraste, numa rede *scale free*, existem grandes variações nos números de ligações por nó. Existe um pequeno número de nós (“centrais”) com muitas ligações coexistindo com um número muito maior de nós com pequeno número de ligações.

Muitas redes do mundo-real que tem sido estudada aparentam ser *scale free*; isto é, elas têm distribuição seguindo uma lei de potência ao invés de uma Gaussiana. Isso é familiar em nossas próprias vidas; existem pessoas que tem um grande grupo de amigos (ou colaboradores, ou fãs), e muito mais pessoas com tal grupo bem menor.

Em 1999 Barabási e Albert propôs o mecanismo de “crescimento por ligação preferencial” como explanação do fato empírico das redes *scale free* [8]. A idéia simplesmente é que os “ricos ficam mais ricos”. Crescimento de redes onde um nó com grande *grau* recebe mais novas ligações do que um nó com *grau* menor. Intuitivamente

isso faz sentido. Pessoas com muitos amigos tendem a conhecer mais novas pessoas e assim fazer mais novos amigos do que pessoas com poucos amigos. Páginas Web referenciadas por muitos apontadores são mais fáceis de achar do que aquelas com poucos apontadores, então novas páginas Web ligam-se àquelas com muitos apontadores.

O modelo de crescimento proposta por Barabási e Albert é definido da seguinte forma: A rede começa com um pequeno número inicial de nós. Em cada passo, um novo nó é adicionado, e é conectado a m nós existentes, para algum m . A probabilidade de fazer uma dessas ligações para um nó i é proporcional ao grau do nó i .

Em vários estudos de redes do mundo-real, incluindo o WWW (World Wide Web), a internet, redes de energia elétrica, redes de conexões aéreas entre cidades, redes colaborativas de cientistas, redes metabólicas em organismos biológicos, redes de proteínas em fermento, e várias outras, Barabási, Albert, e outros descobriram que a distribuição do grau empírica foi bem explicada pela lei de potência $P(k) \sim k^{-\gamma}$, com γ geralmente entre 2 e 3 (dependendo da rede particular) [9].

Barabási e Albert (propagaram os trabalhos anteriores de Price e outros [10]) concluindo que seu simples mecanismo de “crescimento com ligação preferencial” é o que guia a evolução das redes do mundo-real. Barabási afirmou que no mundo-real, “pequenas diferenças em habilidades ou flutuações randômicas podem ter levado a uma grande disparidade com o tempo”.

Também pode ser demonstrado que redes *scale free* possuem as propriedades das redes *small world*. Porém, algumas (não todas) redes *small world* também são redes *scale free*.

A tabela a seguir dá um sumário qualitativo das várias caracterizações das redes descritas acima com respeito aos diferentes modelos de rede.

Modelo da rede	Distribuição do grau	Índice de agrupamento	Distâncias média dos caminhos
Regular	constante	auto	auto
Randômica	Poisson	baixo	baixo
<i>Small world</i> de Watts-	Depende do p	auto	baixo

Strogatz (pequeno e não nulo p)			
<i>Scale free</i> de Barabási-Albert	Lei de potencia	auto	baixo
Resultados empíricos nas redes do mundo-real	Lei de potencia	auto	baixo

Tabela 2 – Sumário qualitativo das caracterizações de redes.

1.3. Busca em redes complexas

Uma grande maioria das aplicações voltadas para redes foca em melhorar os algoritmos de buscas por nós com certas características dentro de uma rede. Isto é em geral um problema difícil. Em grandes redes, o fato de existir um caminho mínimo entre dois nós não implica que seja fácil achá-lo [4].

Usualmente as aplicações envolvendo tais redes são voltadas para resolver problemas localizados, por exemplo, escolher a combinação de ligações que resulta no menor caminho entre dois nós. Outras aplicações buscam encontrar certos nós com propriedades interessantes na rede, por exemplo, pessoas mais influentes numa grande corporação de acordo com a quantidade e para quem são as ligações. Algumas procuram pelos agrupamentos formados, onde os indivíduos dentro dele possuem muitas relações mutuamente. Entendendo tais grupos é possível entender como acontece à troca de idéias e inovações entre grupos, além de mapear aquelas pessoas mais bem informadas ou com várias linhas de pensamentos diferentes.

A aplicação proposta nesse trabalho possui um propósito ainda pouco explorado em termos de redes complexas. Fazendo o papel de grande promessa no campo das inovações colaborativas atuais, as redes sociais têm sido centro das atenções das grandes corporações. Com um público de crescimento espantoso e causando transações comerciais bilionárias, esses sistemas tem sido cada dia mais explorados por serem constituídos, em sua essência, por grande quantidade de informações particulares e ponto central de inúmeros serviços de compartilhamento de informação. Troca de mensagens, blogs, e serviços voltados para gerenciamento de consulta de pessoas. Para seus usuários, as redes sociais tem se tornado ferramenta indispensável para

interatividade com outras pessoas através da Web. A qualidade da rede e sua valorização para certo usuário estão diretamente relacionadas com a quantidade de pessoas com que o usuário se relaciona e com a frequência de interatividade aplicada.

Os sistemas de redes sociais, além de outros serviços, precisam prover ferramentas para consulta e navegação de usuários. As ferramentas de consultas são basicamente filtros para propriedades que compõe os perfis dos usuários da rede. As ferramentas de navegação permitem a visualização da lista de usuários relacionados a cada usuário, dessa forma é possível navegar através das relações e encontrar mais usuários. Essas são, basicamente, as duas formas de acesso quando se pretende adicionar pessoas em sua lista de relacionados. Considerando uma tarefa tediosa e cansativa, quando o único propósito é aumentar o círculo de amizade, navegar através das pessoas procurando por nomes e fotos, torna-se uma atividade pouco realizada fazendo com que a lista de amigos na rede social não reflita a realidade social do usuário. Dessa forma, podemos considerar que numa rede social, que sofre constantes mudanças, existem várias potenciais relações que ainda não foram criadas.

Buscando melhorar a qualidade da rede social como um todo, a ferramenta aqui desenvolvida propõe um mecanismo de busca inteligente, onde os parâmetros utilizados para filtrar e ranquear o resultado depende do contexto. O contexto aqui se refere ao usuário que está buscando. E as pessoas que farão parte da lista resultante possuem propriedades importantes dentro do contexto. Basicamente são analisadas algumas propriedades particulares entre o usuário que está buscando e o possível usuário a sugerir. Em outras palavras, o mecanismo busca identificar àquelas possíveis relações, muitas vezes existentes no mundo real, e ordená-las seguindo alguns parâmetros. Essa busca utiliza a sub-rede na rede social que contém ambos os usuários incluindo alguns outros relacionados. Utilizando algoritmos eficientes o mecanismo realiza cálculos para medir algumas forças de interação entre o usuário buscador e cada um de um grupo pré-selecionado de usuário com grande potencial.

Os algoritmos utilizam apenas a estrutura da rede para mensurar tais forças, deixando de lado dados do perfil dos usuários. A busca fazendo uso dos dados do perfil dos usuários constitui outro tipo de problema, que não será abordado aqui. Como os problemas são independentes, a criação de uma solução envolvendo os dois modelos, considerando a estrutura de relacionamentos e os dados do perfil, provavelmente terá melhores resultados do que cada separado.

As forças, aqui citadas, referem-se a algumas propriedades que podem ser mensuráveis dentro de sub-redes formadas entre os dois nós analisados seguindo algumas regras. Por exemplo, duas pessoas que se conhecem e tem algum tipo de relação provavelmente compartilham de certo conjunto de amigos em comum. Tal grupo pode ser uma turma de escola, companheiros de trabalho ou realmente amigos em comum. Quanto maior o conjunto de amigos em comum entre duas pessoas, maior a probabilidade de elas se conhecerem. Em uma rede social, essa propriedade pode ser medida apenas fazendo a interseção dos conjuntos de nós relacionados de cada usuário envolvido. Além do tamanho do conjunto formado pelos amigos em comum, uma análise para investigar as relações dentro desse conjunto pode fornecer informações que melhore tal força. Um conjunto de amigos compartilhado entre duas pessoas, torna-se mais expressivo se as pessoas dentro desse conjunto se relacionam de alguma forma. O conjunto de pessoas que compartilham relações em comum provavelmente fazem parte de um mesmo agrupamento de pessoas, mecanismos de busca por agrupamentos de nós dentro de uma rede podem explicar tais formações. A descoberta desses agrupamentos pode mostrar a homogeneidade de idéias e filosofias dentro das pessoas que compõe o grupo.

As várias forças exemplificadas acima são medidas relativas a dois pontos na rede, o primeiro, como referência, é o nó da rede que está buscando mais conexões para outros nós, o segundo, um candidato a ser escolhido como novo relacionado ao primeiro. A aplicação do mecanismo de busca proposto dentro de uma rede social, onde a quantidade de usuários pode chegar a 100 milhões, pode se torna inviável computacionalmente se as forças de relacionamento, acima exemplificadas, forem mensuradas partindo de uma pessoa para todas as outras na rede. Porém, uma vez que a rede social for caracterizada como sendo uma rede *small-world*, o processamento dessas mensurações podem ser direcionado para um grupo muito menor de usuários. Fazendo uso do alto valor do coeficiente de agrupamento da rede social, é possível reduzir o conjunto a ser analisado na ordem de milhares. É mais provável que conheças algum amigo dos seus amigos do que qualquer outra pessoa na rede [4]. Dessa maneira, um pré-processamento deve filtrar o conjunto de pretendentes como sendo aqueles que estão ligados a exatamente dois passos de relacionamento com a pessoa que está buscando, excluindo aqueles com muito baixa ou nenhuma probabilidade de ser escolhida no resultado final.

1.3.1. Aplicabilidade

A proposta desse projeto é a criação de um mecanismo de sugestão de novos relacionamentos para certo usuário de uma rede social, porém, os conceitos aqui introduzidos podem ser aplicados em muitas outras redes que apresentem características de uma rede *small-world*. As forças que unem dois usuários dentro de uma rede social podem ser utilizadas com outros propósitos. Por exemplo, dentro de uma rede de computadores, como a internet, o mecanismo de forças pode ajudar na tomada de decisão para direcionar investimentos em infra-estrutura conectando roteadores específicos. Outras aplicações podem utilizar o mecanismo para sugestão de produtos. Nesse caso, a dificuldade maior encontra-se em definir como funciona a rede, uma vez que não, necessariamente, existe uma rede social por trás. Uma rede social virtual pode ser extraída de uma base de informações de clientes e produtos comprados por eles. Cada nó da rede continuaria sendo uma pessoa, um cliente no caso, e as relações entre elas seriam extraídas da lista de produtos comprados. Duas pessoas estão ligadas se compraram o mesmo produto ou semelhante. Com essa rede construída, podem ser feitas análises para encontrar padrões de compras e sugestões de produtos podem ser feitas baseadas nas forças entre dois clientes. Um sistema de sugestão como esse pode ser modelado, por exemplo, para uma rede de locadora de filmes, onde existem cadastros de clientes e históricos de aluguel de filmes antigos. Os filmes poderiam ser classificados por gênero, atores, época, etc. para construção da rede de compras.

A internet e as intranets estão repletas de sistemas de cadastros de clientes e lista de produtos comprados. Muitos sistemas já utilizam essas bases para fazerem sugestão de novos produtos, porém, a abordagem geralmente utilizada está voltada para comparações de perfis de compra. Análises sobre redes complexas construídas com essas informações podem revelar relações que dificilmente seriam reveladas em outras abordagens. Entender como funcionam os agrupamentos de clientes que seguem alguns padrões de compra pode ajudar na previsão e exploração de certas linhas de produtos para esses clientes.

2. DESENVOLVIMENTO

O desenvolvimento deste trabalho apresenta além de fundamentos teóricos, muita produção prática. Para que os resultados planejados inicialmente fossem alcançados foi necessária a produção de um framework capaz de criar o ambiente de uma rede social, além de possibilitar a execução de algoritmos eficientes para o procedimento de caracterização da rede e os de recomendação.

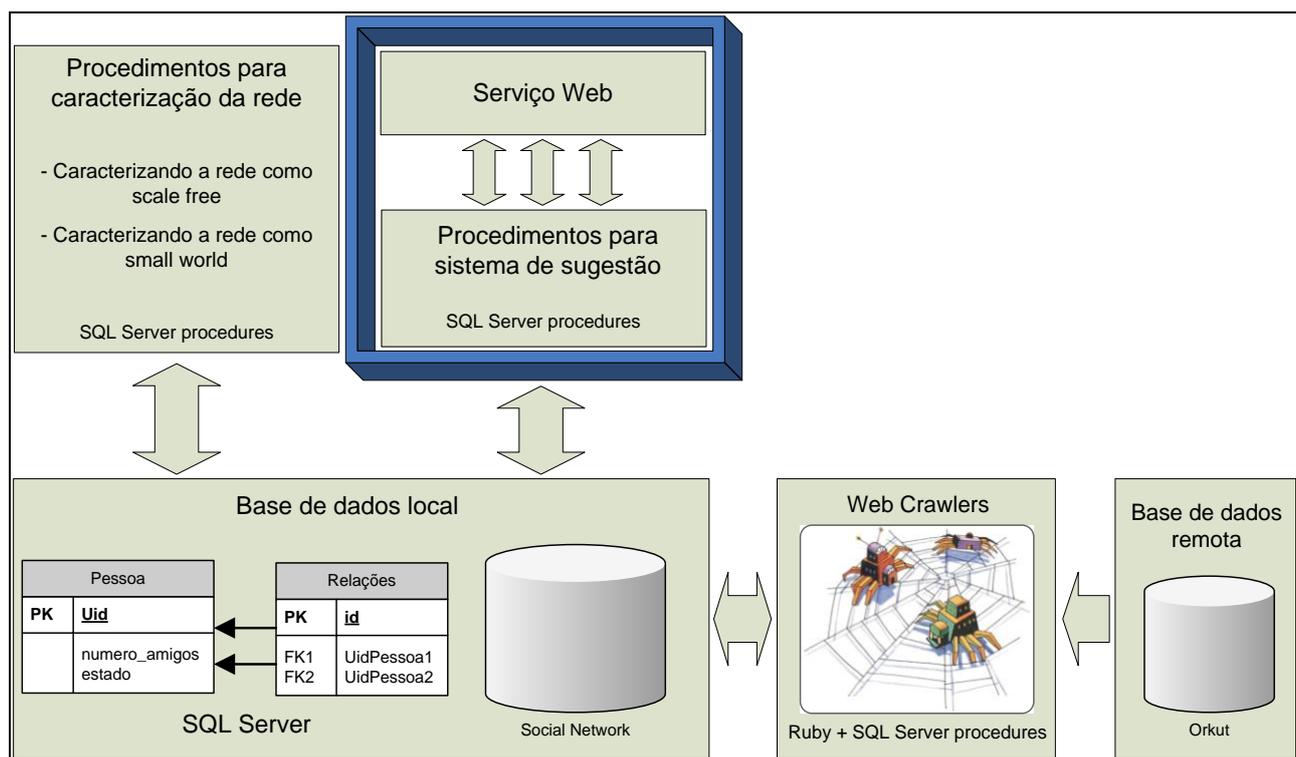


Figura 4. Macro arquitetura do framework.

O framework é composto dos seguintes módulos:

- **Base de dados local:** Esquema de tabelas em uma base de dados local onde estão as informações que constituem a estrutura da rede social. A utilização de um banco de dados local é extremamente essencial para organizar os

dados e prover uma estrutura para execução dos procedimentos de grande custo computacional de maneira eficiente.

- **Web crawler:** Subsistema desenvolvido com a linguagem de programação *ruby* e que é responsável por popular a base de dados local com os dados extraídos de uma rede social real. A rede social utilizada é o Orkut, rede mantida pela corporação Google. O controle da ordenação das páginas para extração da rede social foi concebido de maneira a buscar os dados essenciais para o processo de recomendação de uma única pessoa. O framework é um ambiente para testes, todavia o funcionamento real da aplicação de recomendação precisa de acesso dedicado à base de dados real da rede social. O ideal realmente seria que a aplicação funcionasse dentro do próprio sistema da rede social, maximizando o desempenho do serviço.
- **Procedimentos para caracterização da rede:** Este módulo é composto por procedimentos escritos puramente em T-SQL (a linguagem de script SQL do SQL Server) que implementam os algoritmos utilizados para caracterização da rede social como sendo redes *small world* e *scale free*. A caracterização é feita comprovando a existência de algumas propriedades através de mensurações matemáticas sobre o grafo formado pela rede social.
- **Módulo do serviço de recomendação:** Este módulo contém os procedimentos, também em T-SQL, utilizados para calcular as métricas, anteriormente citadas como *forças* entre pontos da rede, utilizadas no processo de recomendação. Para que a solução seja completa, a confecção de um serviço web, para utilização geral do sistema, é imprescindível. Por questões de escopo esse módulo será desenvolvido em breve.

2.1. Construindo a rede social

Para que os procedimentos aqui desenvolvidos pudessem rodar com máximo desempenho foi necessária a construção de uma base de dados robusta. O banco de dados utilizado é o SQL Server 2005 da Microsoft. A base de dados criada contém basicamente duas tabelas, *person* contendo os registros que representam cada usuário (nós do grafo) e *personPerson* contendo relacionamentos entre os usuários (as arestas do grafo). Todos os algoritmos implementados nesse projeto foram escritos em T-SQL,

uma linguagem não orientada a objetos, dessa forma alguns outros campos foram introduzidos na tabela *person* para armazenar dados de controle de estado para vários dos procedimentos utilizados.

O processo de extração dos dados estruturais da rede social é feito através de web crawlers programados de forma automática para navegar na rede social, seguindo uma ordem, e armazenar dados de cada usuário assim como a lista de pessoas com quem esse usuário se relaciona. Imaginando um grafo como uma estrutura em árvore, com alguns loops, a ordem de visita do crawler segue uma busca em largura.

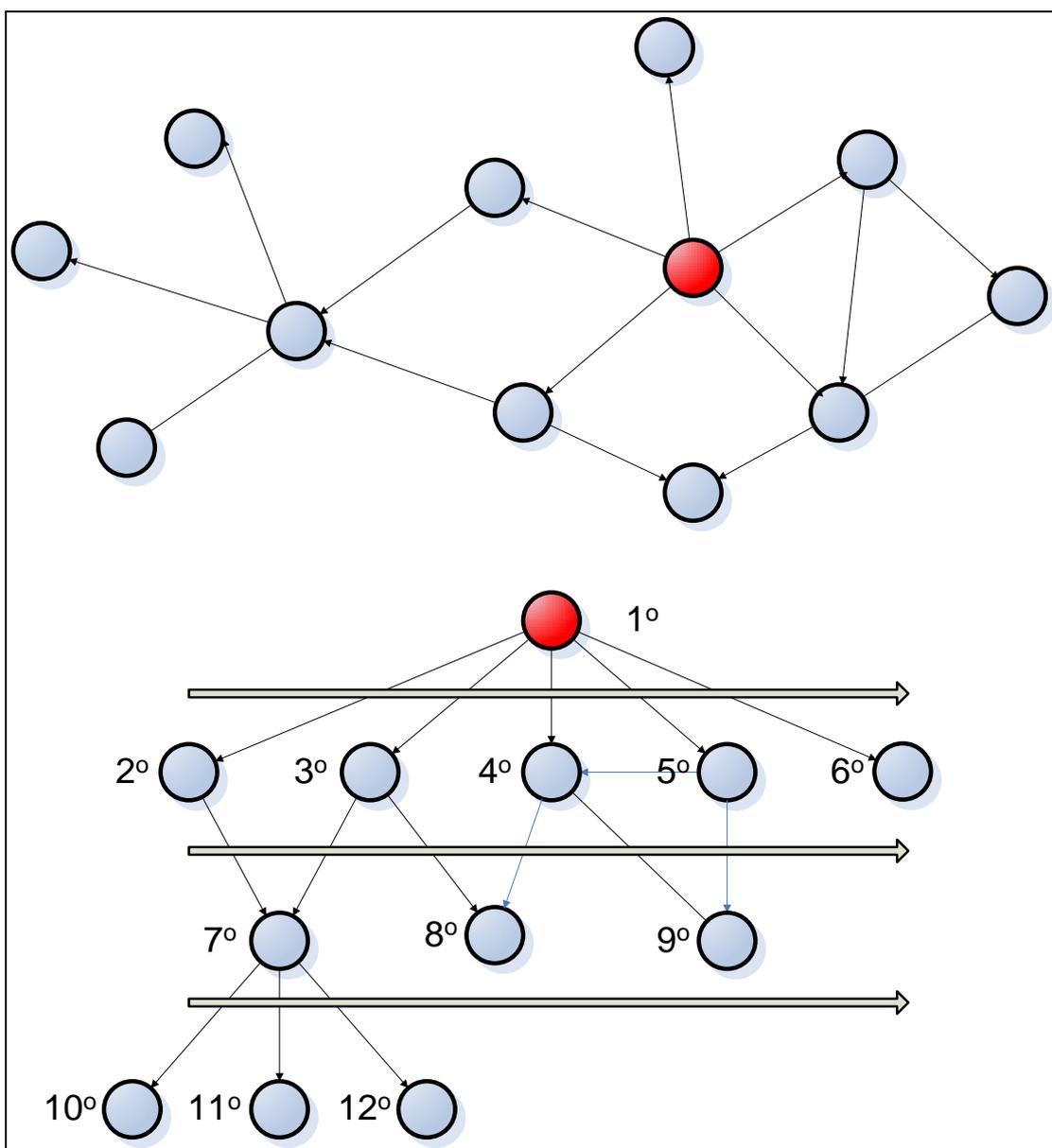


Figura 5. Imagens ilustrativas da rede social (superior) e da visão em árvore da mesma (inferior).

Na prática o processo de mapear a rede social inicia-se a partir de um nó raiz (nó vermelho na figura 5). Uma instancia do web crawler visita a página inicial daquele

usuário, em seguida extrai algumas informações pessoais desse usuário, o passo a seguir é uma repetição do procedimento a seguir. O web crawler visita a página com a lista de amigos daquele usuário, extrai e armazena a identificação de cada um deles em uma lista de usuário para visitar. A identificação desse usuário será utilizada para que outra instância do web crawler faça a visita à página da rede social desse outro usuário. Após esse processo, o web crawler busca uma identificação de usuário que ainda não foi visitado e que tenha esperado pela visita por mais tempo. Dessa forma o processo, como um todo, acontece como uma busca em largura, buscando primeiramente os usuários relacionados com o usuário raiz, em seguida busca pelos amigos desses últimos e assim por diante.

A rede social Orkut possui mais de 70 milhões de usuários, muito grande para ser mapeada através de crawlers na web. Por tais circunstâncias, o processo de criação da rede social local foi minimizado para mapear duas sub-redes, uma partindo de um usuário brasileiro e outra partindo de um usuário indiano. A decisão de escolher um indiano como segunda semente para o processo de extração é que o público indiano é o que cresce mais rápido perdendo em quantidade para os brasileiros e, por muito pouco, para os americanos.

2.2. Caracterização

Nesta seção são detalhados os processos de caracterização da rede social em análise como sendo uma rede *scale free* e também *small world*. Os dois modelos de redes foram apresentados na seção abordando redes complexas. A caracterização da rede social se faz necessária uma vez que é essencial para o entendimento e aceitação de alguns dos passos importantes do processo de recomendação. Entendendo o conceito das redes *small world* e o funcionamento dos agrupamentos com seus coeficientes, é essencial para viabilizar a etapa de pré-processamento para filtrar a rede, além de inspirar os algoritmos na etapa de ponderação.

A caracterização é feita através da realização de alguns cálculos e constatação de algumas propriedades conceituais dos modelos de redes *scale free* e *small world*. Como foi esclarecido no início deste capítulo, todos os dados sobre a rede social foram armazenados em um banco de dados, e os procedimentos para caracterização escritos em linguagem de script T-SQL. Os procedimentos a seguir foram executados sobre duas bases de dados distintas. Uma que foi preenchida com dados de um pedaço da rede

social Orkut com usuários brasileiros, a segunda base foi preenchida com um pedaço da mesma rede, porém com dados de usuários indianos.

2.2.1. Caracterizando a rede como *scale free*

Como foram apresentadas no capítulo sobre redes complexas, as redes *scale free* se diferenciam das redes de formação randômica pelo fato de apresentarem grandes variações no grau dos nós. O grau de um nó refere-se à quantidade de nós ligados a ele. Além disso, a distribuição do grau dos nós de uma rede *scale free* segue uma função de lei de potência $P(k) \sim k^{-\gamma}$.

Da forma que foram armazenados os dados, não são necessários nenhum cálculo prévio sobre eles para constatação dessa propriedade, apenas uma consulta com algumas funções agregadoras para contagem de nós relacionados e sua frequência. Os resultados foram os seguintes:

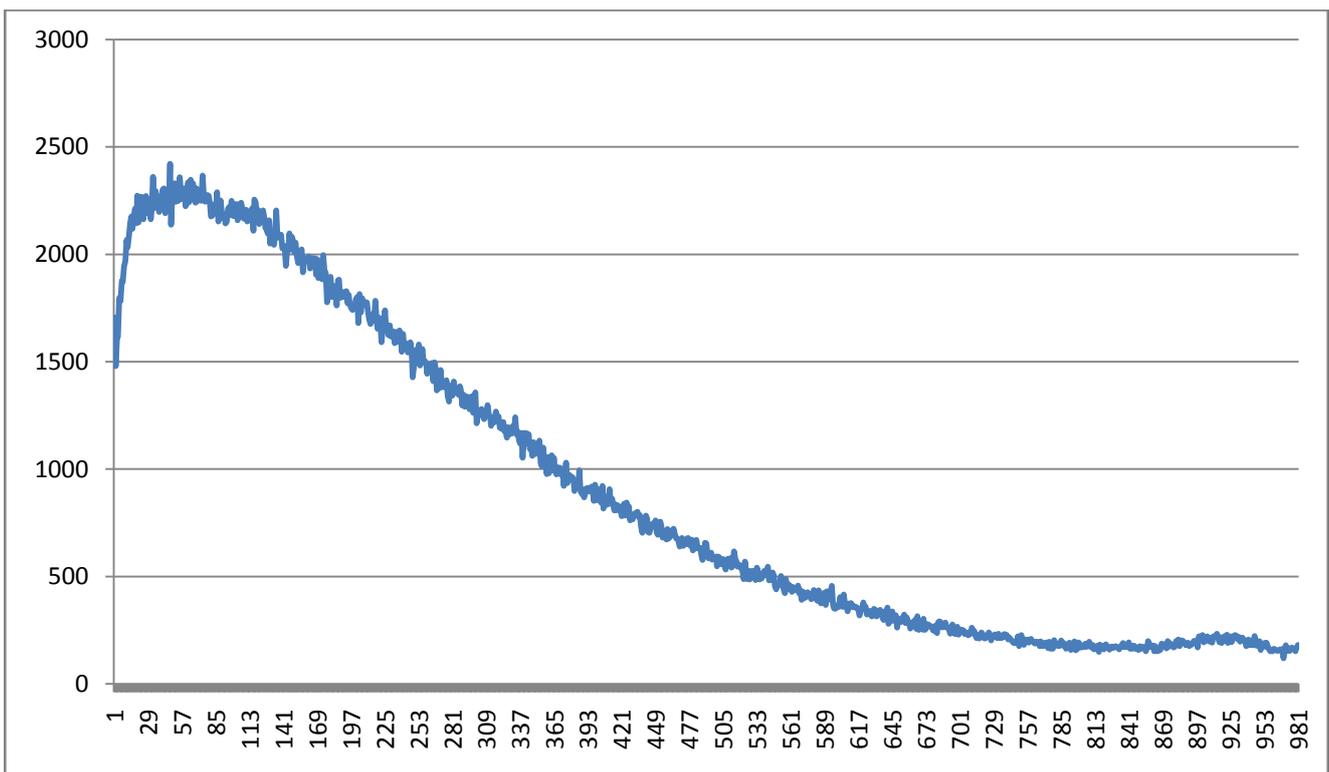


Figura 6 – Gráfico de distribuição do grau dos nós da rede social Orkut com dados de usuários brasileiros.

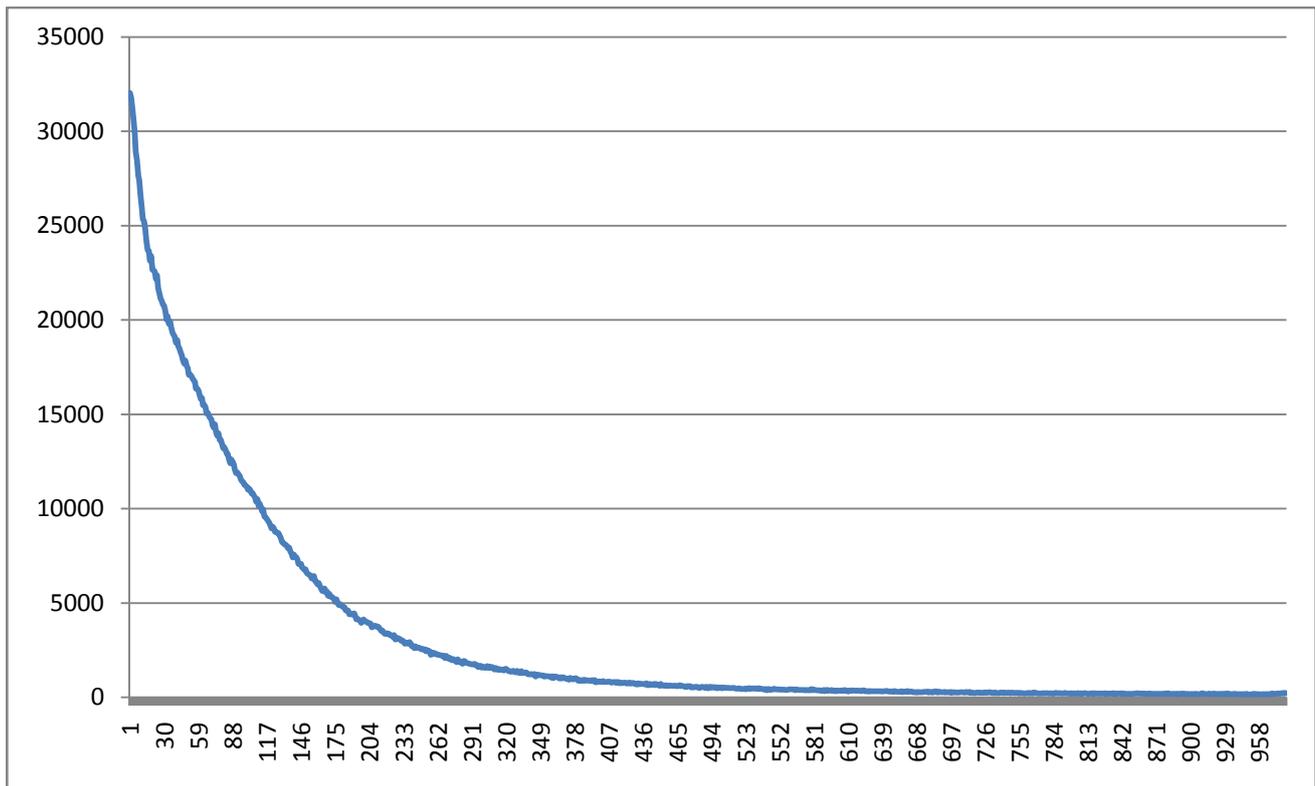


Figura 7 – Gráfico de distribuição do grau dos nós da rede social Orkut com dados de usuários indianos.

O gráfico na figura 6 representa a distribuição do grau dos nós da rede social formada por dados dos usuários brasileiros, a figura 7 representa o mesmo gráfico, porém com dados de usuários indianos. Para geração do gráfico da figura 6 foram utilizadas **893.295** contas pessoais extraídas de uma sub-rede do Orkut dos usuários brasileiros. Para o gráfico da figura 7 foram utilizadas **3.133.267** contas pessoais extraídas de uma sub-rede do Orkut dos usuários indianos.

Com exatidão é fácil afirmar que ambos os gráficos seguem a lei de potência esperada em redes *scale free*. O segundo gráfico segue uma melhor uniformidade em relação ao primeiro. Isto pode ser explicado como consequência da estratégia de extração delegada aos crawlers na etapa de construção da rede social, e também pela quantidade de dados utilizados em cada gráfico. A estrutura de grafo formada por uma rede social pode não apresentar uniformidade em toda sua extensão. As propriedades, como a média dos graus de cada nó, podem mudar. Nós com maior grau estão localizados numa região mais densa e mais interna, especialmente falando. Dessa maneira, uma análise feita a partir de um sub-grafo de uma estrutura maior pode exibir deformações, uma vez que o processo de extração dos dados não tenha alcançado algumas regiões, a borda ou periferia, por exemplo, onde existem nós com média de grau menor. É importante definir a utilização do termo “alcançar” uma região do grafo,

uma vez que redes *small world* apresentam a média dos menores caminhos entre pares de nó muito pequeno. A navegabilidade através dos nós é muito lenta, pois a estrutura apresenta um diâmetro muito pequeno.

Outra característica a ser comparada entre os dois gráficos é a média do grau dos nós envolvidos. Para a rede indiana a média do grau dos **750.000** primeiros nós extraídos é de **140**, já para a rede brasileira a média com a mesma quantidade de nós sobe para **275**. Um pouco estranho a primeira vista, porém realista em certas perspectivas. A grande diferença entre os dois valores mostra como a média do grau dos nós pode variar. Essa variação pode levantar algumas hipóteses, a idade da rede, ou sub-rede, pode ser considerada. Historicamente os brasileiros utilizam o Orkut mais e há mais tempo que muitos outros países, considerando o público em média. Isso leva-nos a supor algumas outras coisas sobre a dinâmica de uma rede social, abrindo caminho para um estudo minucioso das causas e impactos na velocidade de crescimento.

Procedimentos para comparações entre várias sub-redes de um grande grafo, como uma rede social, necessitam de acesso direto aos dados, impossibilitando a utilização do mesmo framework baseado numa extração através de uma interface Web.

Outra hipótese seria considerar fatores culturais localizados que influenciam na média real de relacionamentos por pessoa. Todas essas hipóteses são apenas conjecturas ficando como trabalhos futuros o estudo aprofundado para devidas provas e contestações.

[4] Toda rede *scale free* apresentam também as propriedades das redes *small world*. O próximo tópico foi reservado para os procedimentos de caracterização da rede social como *small world*.

2.2.2. Caracterizando a rede como *small-world*

As redes *small world* são caracterizadas a partir da constatação de duas propriedades distintas: o alto valor do coeficiente de agrupamento médio e a baixa média dos menores caminhos entre pares de nós.

2.2.2.1. Coeficiente de agrupamento

O coeficiente de agrupamento é definido como a probabilidade de que dois nós distintos e relacionados a outro mesmo nó estejam relacionados entre si. Em outras palavras mede a probabilidade de que dois amigos comuns de uma pessoa sejam amigos

entre si. Como esclarecido por Melanie Mitchell, “é mais provável que você conheça um amigo do seu amigo do que qualquer outra pessoa aleatoriamente” [1].

Primeiro vamos definir um grafo em termos de um conjunto de n vértices $V = v_1, v_2, \dots, v_n$ e um conjunto de arestas E , onde e_{ij} denota uma aresta entre vértices v_i e v_j . Abaixo assumimos que v_i, v_j e v_k são membros de V [6].

Nós definimos de vizinhança N para um vértice v_i como seus vizinhos conectados imediatamente como a seguir:

$$N_i = \{v_j\} : e_{ij} \vee e_{ji} \in E$$

O grau k_i de um vértice é o número de vértices, $|N_i|$, em sua vizinhança N_i .

O coeficiente de agrupamento C_i para um vértice v_i é a proporção de ligações entre os vértices que estão em sua vizinhança dividida pelo número de ligações que poderiam existir entre eles. Para um grafo direcionado, e_{ij} é distinto de e_{ji} , e então para cada vizinhança N_i existem $k_i(k_i - 1)$ ligações que podem existir dentre os vértices na vizinhança (k_i é o grau total (entrada + saídas) de um vértice). Então, coeficiente de agrupamento para um grafo direcionado é dado como:

$$C_i = \frac{|\{e_{jk}\}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{jk} \in E.$$

Um grafo não direcionado tem a propriedade de que e_{ij} e e_{ji} são considerados idênticos. Então, se um vértice v_i tem k_i vizinhos, $\frac{k_i(k_i-1)}{2}$ arestas podem existir dentre os vértices na vizinhança. Então, o coeficiente de agrupamento para um gráfico não direcionado pode ser definido como:

$$C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{jk} \in E.$$

O coeficiente de agrupamento para todo o sistema é dado por Watts e Strogatz como a média do coeficiente de agrupamento de cada vértice:

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$$

A implementação do procedimento que calcula o coeficiente de agrupamento para cada vértice na rede está escrito em T-SQL na figura a seguir:

```

--Esta tabela armazena todos os relacionamentos entre meus amigos (diretos)
mutuamente

declare @RelashionTable table(Uid1 numeric(20), Uid2 numeric(20))
INSERT @RelashionTable
    SELECT pp.SourceUidFrom, pp.SourceUidTo
        FROM dbo.PersonPerson pp
        JOIN (
            SELECT pp1.SourceUidTo as Uid1, pp2.SourceUidTo as Uid2
            FROM dbo.PersonPerson pp1
                CROSS JOIN
                dbo.PersonPerson pp2
            WHERE pp1.SourceUidFrom = @PersonVisiting
                and pp2.SourceUidFrom = @PersonVisiting
        ) ppPossible on ppPossible.Uid1 = pp.SourceUidFrom
            and ppPossible.Uid2 = pp.SourceUidTo

--Agora conto quantos amigos diretos possuo evitando os registros duplicados
--mesmo na direção inversa do relacionamento

INSERT @RelashionTable
    SELECT Uid2, Uid1 FROM @RelashionTable

declare @RelashionTable table(Uid1 numeric(20), Uid2 numeric(20))
INSERT @RelashionTableUnique SELECT DISTINCT Uid1, Uid2 FROM @RelashionTable
--Agora tenho dois e somente duas cópias do mesmo registro, um em cada direção
--Conto quantos são os possíveis relacionamentos entre todos meus amigos

SELECT @FriendsCrossRelashion = COUNT(id) FROM dbo.PersonPerson
        WHERE SourceUidFrom = @PersonVisiting

SET @FriendsCrossRelashion =
    ((@FriendsCrossRelashion * (@FriendsCrossRelashion - 1)) / 2)

-- A probabilidade de dois amigos meus se conhecerem é igual à quantidade que
--existe realmente dividido pelas possíveis relações

SELECT @RelashionProbability = (
    CAST((COUNT(Uid1)/2) AS decimal(18,8))
    /
    CAST(@FriendsCrossRelashion AS decimal(18,8))
    )
    FROM @RelashionTableUnique

```

Figura 7 – Gráfico de distribuição do grau dos nós da rede social Orkut com dados de usuários indianos.

O algoritmo é executado para cada nó da rede. O coeficiente de agrupamento da rede é a média dos coeficientes de agrupamentos de cada nó. Para a sub-rede com os

dados dos usuários brasileiros o valor do coeficiente de agrupamento é igual a **0.1830**, já para a sub-rede com os dados dos usuários indianos esse valor sobe para **0.3478**.

É indispensável uma análise dos resultados. Apesar de a rede indiana apresentar menor média do grau dos nós, ela apresenta maior coeficiente de agrupamento. O coeficiente **0.3478** mostra que em média, dentro de cada grupo dos possíveis relacionamentos, dois a dois, em média 34,78% existem. Por exemplo, se esse valor fosse 40%, num grupo de 5 pessoas, existem 4 dos 10 possíveis relacionamentos.

Em comparação, a rede indiana apresenta agrupamentos menores, porém mais isolados e coesos, do que a rede brasileira. Considerando que uma rede social virtual está em constantes mudanças, este coeficiente pode ser utilizado para medir se a sub-rede está próxima de se estabilizar ou não. Fazendo uma comparação com o coeficiente da sub-rede com o coeficiente de uma sub-rede maior que engloba a primeira podemos saber se a tendência levará aquela sub-rede menor a aumentar ou não esse coeficiente.

A utilização do mecanismo de sugestão proposto nesse projeto, matematicamente, tem o propósito de aumentar esse coeficiente para certo usuário, uma vez que busca agregar à sub-rede do usuário, conexões para nós onde o conjunto de amigos em comum já possui grande coeficiente.

Para que uma rede seja caracterizada como *small world* é necessário que o coeficiente de agrupamento seja significativamente maior do que o coeficiente de agrupamento de uma rede randômica construída com o mesmo conjunto de vértices, além de apresentar um pequeno valor para a média dos menores caminhos entre os pares de nós. O cálculo da média dos menores caminhos entre pares de nós é feito a seguir.

2.2.2.2. Caminho mínimo médio entre pares de nós

O número de arestas em um caminho que conecta os vértices i e j é chamado de distância do caminho [11]. O caminho geodésico (ou caminho mínimo), entre os vértices i e j , é um que conecte esses vértices com distância mínima (podem existir vários caminhos geodésicos entre dois vértices); a largura do caminho geodésico é a distância geodésica d_{ij} entre os vértices i e j . A distância geodésica média é definida como:

$$\ell = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}.$$

Onde N é o tamanho do conjunto dos vértices do grafo.

A tarefa de achar um caminho geodésico é um problema clássico estudado por vários autores que desenvolveram seus próprios algoritmos. O mais comum desenvolvido por Dijkstra, onde a complexidade temporal é $O(N^2)$. Para um grafo com 3 milhões de nós torna-se ineficiente a utilização de Dijkstra. Para viabilizar a constatação dessa propriedade no grafo foi necessário desenvolver um método de uso mais específico e eficiência com complexidade temporal menor. Para tal, foi criado um algoritmo baseado no método de Dijkstra. Utilizando técnicas de programação de dinâmica, essa variância de Dijkstra permite reutilizar as informações intermediárias que seriam perdidas no método usual. Dessa maneira, o cálculo das distâncias de um nó para todos os outros pode ter a complexidade temporal reduzida para $O(N \log N)$.

A seguir, uma explicação do algoritmo desenvolvido utilizando uma analogia com um grupo de pessoas. Suponhamos que dentro de uma sala de aula existam 100 alunos distribuídos de forma aleatória. A distância entre duas pessoas nessa sala é igual a quantidade de pessoas que precisam repetir uma mensagem para que ela seja transmitida de uma pessoa para a outra, considerando que a repetição da mensagem é feita apenas entre pessoas localizadas muito próximas. A regra de transmissão de mensagem é: ela deve ser transmitida para todas as pessoas próximas excluindo quem a transmitiu anteriormente. A mensagem é um número, e ao transmiti-lo deve-se incrementar seu valor em “1”. Dessa maneira, para calcular a distância para todas as outras pessoas partindo de uma pessoa j qualquer faça o seguinte procedimento:

A pessoa j inicia guardando para si a mensagem “0”. Até que todos tenham um número, execute a transmissão da mensagem seguindo as regras de propagação. As pessoas mais próximas da pessoa j receberam a mensagem “1”. Cada pessoa que recebeu a mensagem precisa propagá-la. As pessoas mais próximas daquelas que possuem “1”, excluindo a pessoa j , receberam a mensagem “2”, e assim sucessivamente. No final todas as pessoas possuem alguma mensagem, o número da mensagem é igual a quantidade de pessoas que a transmitiram partindo da pessoa j . A média dos valores armazenados por cada pessoa é igual à média das distâncias geodésicas partindo da pessoa j até todas as outras na sala. Para o mesmo cálculo para a pessoa k todos devem esquecer a mensagem que possuem e a pessoa k deve iniciar novamente a propagação da mensagem com “0”. O procedimento é feito para cada pessoa na sala, no final, todos possuem a distância geodésica média para todos os outros na sala.

No artigo “*Topology and Shortest Path Length Evolution of The Internet Autonomous Systems Interconnectivity*” [12] é construído um algoritmo também com o propósito de calcular a média de todos os caminhos geodésicos de um grafo. O algoritmo, chamado de “método de Friburgo” utiliza a mesma estratégia descrita acima. O que a primeira vista pareceu uma coincidência agora vêm reafirmar a eficiência da solução encontrada.

A figura a seguir ilustra o grafo em formato de árvore com os valores propagados a partir de um nó raiz:

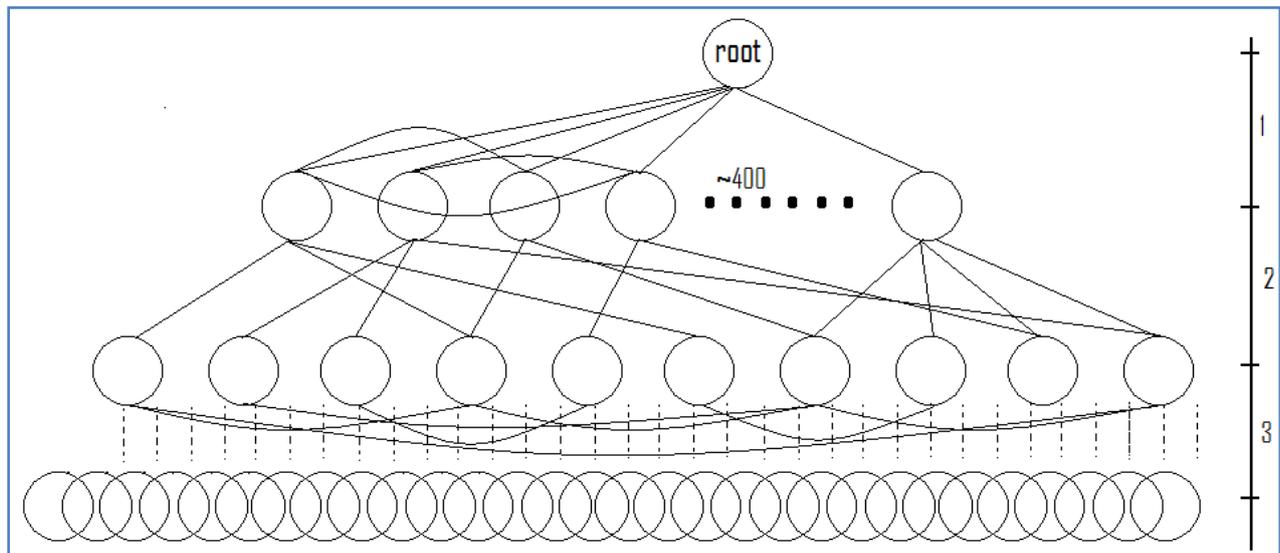


Figura 8 – Estrutura de um grafo em forma de árvore. A barra do lado direito mostra a mensagem transmitida de um nível ao nível inferior. Na rede social analisada a quantidade de relacionamentos de uma pessoa pode variar de 1 a 1000.

Os procedimentos a seguir são a implementação do método proposto utilizando a linguagem de script T-SQL:

```

--Plante-se a semente

update dbo.Person set height = 0 where uid = @RootPerson

--Enquanto existirem pessoas não visitadas faça o seguinte

while exists(select top 1 uid from dbo.Person where height is null and state =
'visited') and @pare = 0 begin
    --Atualize a altura na árvore das pessoas que não tem altura e que
    --conhecem alguém que tem altura com o valor dele mais 1

    update p2 set p2.height = p1.height + 1
    FROM dbo.Person p1
    join dbo.PersonPerson pp on pp.sourceUidFrom = p1.uid
    join dbo.Person p2 on p2.uid = pp.sourceUidTo
    WHERE p1.height is not null
    and p2.state = 'visited' and p2.height is null

```

```

--Caso especial quando pode existir uma ilha fechada no grafo o que faz
esse procedimento não convergir
IF @@ROWCOUNT = 0 begin
    set @pare = 1
end
end
end

--Atualiza o rootPerson com a média das distâncias calculada
UPDATE dbo.Person set avgPathLength = (SELECT AVG(CAST(Height AS decimal(5,3)))
FROM dbo.Person)
WHERE uid = @RootPerson

```

Figura 9 – Procedimento em T-SQL utilizado para calcular a distância geodésica média de um usuário para todos os outros.

```

--Selecionando os nós que precisam receber a distância geodésica média

DECLARE person_cur CURSOR STATIC FOR
    SELECT uid FROM dbo.Person WHERE state = 'visited' and avgPathLength is null
OPEN person_cur
FETCH NEXT FROM person_cur INTO @RootPerson

WHILE @@fetch_status = 0 begin
    --Seta a altura de todas as pessoas para null
    UPDATE dbo.Person set Height = null WHERE Height is not null
    -- Calcula a distância da pessoa rootPerson para todas as outras usando a
    --procedure abaixo

    exec dbo.pr_CountHeight @RootPerson = @RootPerson
    -- Atualiza o rootPerson com a média das distâncias calculada

    UPDATE  dbo.Person  set  avgPathLength  =  (SELECT  AVG(CAST(Height  AS
decimal(5,3))) FROM dbo.Person) WHERE uid = @RootPerson

```

Figura 10 – Este procedimento é a chamada do procedimento anterior repetidamente para cada nó da rede.

Para a sub-rede social com os usuários brasileiros que contém **1.935.003** pessoas, a distância geodésica média é igual a **3,1454**. Para a sub-rede social com os usuários indianos que contém **3.133.267** pessoas, essa mesma métrica é igual a **3,7722**. Isso significa que em média, duas pessoas estão separadas na rede por **3.7** relacionamentos, ou em outras palavras como exemplifiquei anteriormente, a mensagem em média alcançou o valor “**3.7**”.

Para um aumento de 62% na quantidade de nós na sub-rede social indiana em relação à sub-rede social brasileira, há um aumento de apenas **0,6268** na distância

geodésica média. Isso é um aumento logarítmico, comportamento esperado em redes *small world*.

2.3. Mecanismo de sugestão

O mecanismo de sugestão é um procedimento que utiliza os dados estruturais da rede para filtrar e ordenar um conjunto de vértices que possuem algumas propriedades importantes em relação a um dado vértice v_i . Os vértices do conjunto resultante são sugestões para novas arestas ligando ao vértice v_i . A criação dessas novas arestas tem o propósito de melhorar algumas propriedades do vértice v_i , além de prover benefício para a rede como um todo.

O mecanismo de sugestão foi, inicialmente, projetado para sugerir novos relacionamentos para um usuário de rede social. A qualidade da rede social depende colaborativamente de todos. Em quase sua totalidade, as novas relações são criadas após um processo de busca por usuários, utilizando os serviços da rede social para navegar através dos relacionamentos existentes. O propósito de um mecanismo de sugestão é evitar a etapa de navegação e busca por usuários. Com essa facilidade os usuários aumentam mais rapidamente sua lista de relacionamentos, o que possibilita maior interatividade na rede. A qualidade do resultado do mecanismo para um usuário depende diretamente da qualidade dos relacionamentos dos usuários já relacionados diretamente com ele. Nessa perspectiva fica clara a capacidade colaborativa do mecanismo.

O processo de sugestão é dividido em duas etapas: uma filtragem seguida de uma ordenação. A filtragem é uma etapa importante, pois separa os nós com maior possibilidade reduzindo muitas vezes a quantidade total de vértices da rede. A etapa de ordenação considera algumas propriedades interessantes para colocar os vértices mais relevantes no começo da lista resultante.

2.3.1. Filtrar

A etapa de filtragem tem um papel extremamente importante. Ela é responsável por limitar a quantidade de vértices no conjunto que será ordenado na etapa seguinte. Isso é necessário para viabilizar o mecanismo de sugestão uma vez que reduz muito o custo computacional total do processo.

O algoritmo que rege a regra de filtragem faz uso do conceito da propriedade *clustering coefficient* que caracteriza as redes *small world*. Como citado anteriormente: “é mais provável que você conheça um amigo do seu amigo do que qualquer outra pessoa aleatoriamente” [1]. Seguindo essa regra, a etapa de filtragem se restringe a selecionar os vértices adjacentes aos vértices adjacentes ao vértice central do processo de sugestão.

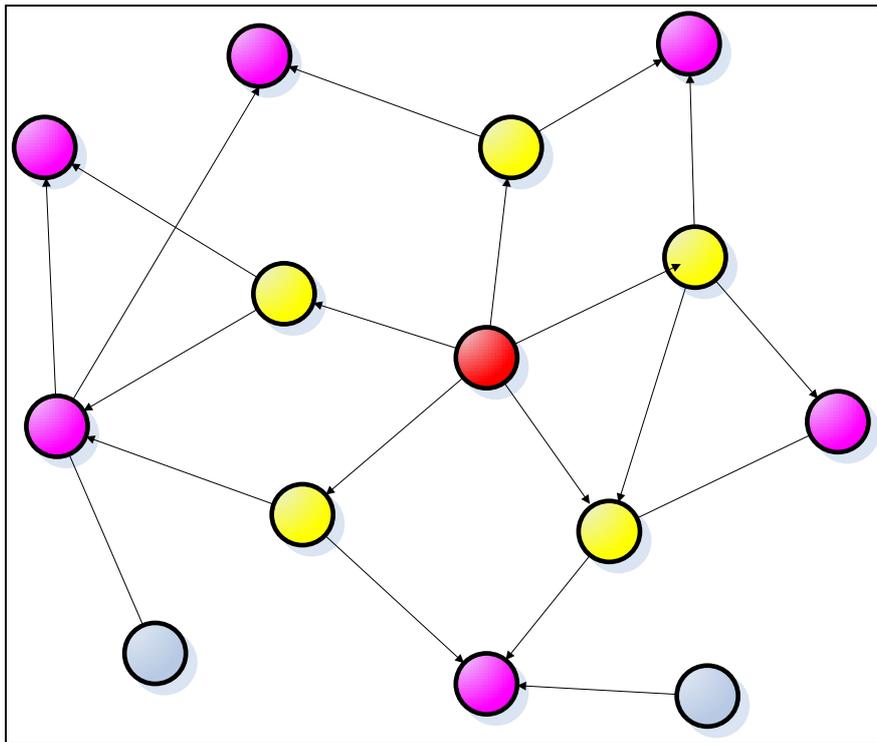


Figura 11 – O vértice em vermelho é o elemento central. Os vértices rosados são os selecionados pela etapa de filtragem.

Na figura acima o vértice vermelho é o elemento central do processo total de sugestão. Os vértices amarelos estão conectados diretamente com o vértice vermelho. Os vértices rosados são os selecionados pela etapa de filtragem. Em média, para os dois casos experimentais, o conjunto dos vértices selecionados pela etapa de filtragem contém 70 mil vértices.

2.3.2. Ordenar

A etapa de ordenação é um pouco mais complexa. Ela consiste na mensuração de alguns índices e a ponderação deles utilizando um mecanismo de auto-ajuste. Não sendo diferente de um mecanismo de ordenação comum, esse também utiliza um único valor numérico relacionado a cada vértice a ser ordenado. Esse valor de indexação pertencente a cada vértice é resultado de um processo que mede a força de interação

entre esse vértice e o vértice central de todo o processo de sugestão. O mecanismo de auto-ajuste regula os pesos que ponderam os índices baseado nos vértices já relacionados ao vértice central do processo de sugestão.

2.3.2.1. Calculando os índices

A mensuração dessa força de interação é resultado de uma média ponderada entre 3 índices independentes. Eles medem propriedades específicas de um sub grafo formado pelos dois vértices em análise e juntamente com outros que possuem alguma relação com os mesmos. São explorados 3 índices por razões maiores de tempo e escopo do projeto, porém quanto mais índices melhor para que a precisão na ponderação seja maior. O processo de ponderação, ou seja, de escolher pesos para os índices, será mais bem discutido na próxima seção, A seguir uma formalização sucinta de cada um dos índices.

Algumas definições úteis na formalização dos índices:

Defina C_i como sendo o conjunto dos vértices que são adjacentes ao vértice v_i . Matematicamente se resume aos vértices da lista de adjacência do vértice v_i .

Defina como D_C a densidade de adjacência entre os vértices contidos no conjunto C . A densidade é calculada como a quantidade real das arestas onde os vértices de origem e destinos estão em C dividida pela quantidade de arestas do *clique* (grafo completo) formado por todos os vértices contidos em C .

$$D_C = \frac{\sum_{i \in C} (\sum_{j \in C} (M_{ij}))}{(|C| * (|C| - 1)) / 2}$$

Onde M_{ij} é o elemento ij da matriz de adjacência.

1. O primeiro índice é definido como a quantidade de vértices adjacentes, ao mesmo tempo, ao vértice i e ao vértice j , onde i é o vértice centro da análise e j é o vértice que está sendo ordenado para a sugestão. Dessa forma temos:

$$I_{1ij} = |C_i \cap C_j|$$

Fazendo uma analogia com redes sociais, esse índice mede a quantidade de amigos em comum da pessoa i e da pessoa j .

A implementação em T-SQL do calculo do índice 1 é descrito a seguir:

```
--quantidade de relacionamentos em comum entre os dois
Select @retorno = COUNT(pp1.SourceUidTo) FROM dbo.PersonPerson pp1
                JOIN dbo.PersonPerson pp2 ON pp2.SourceUidTo = pp1.SourceUidTo
WHERE pp1.SourceUidFrom = @uid1
      AND pp2.SourceUidFrom = @uid2
```

Figura 12 – Implementação do algoritmo do índice I_1 em T-SQL.

2. O segundo índice se refere à densidade do conjunto mensurado no primeiro índice:

$$I_{2ij} = D_{C_i \cap C_j}$$

Fazendo uma analogia com a rede social analisada, esse índice mede o nível de coesão dentro do “pequeno” grupo formado pelos amigos em comum da pessoa i e da pessoa j . Se esse índice tiver um valor pequeno então as pessoas desse grupo não se relacionam bem.

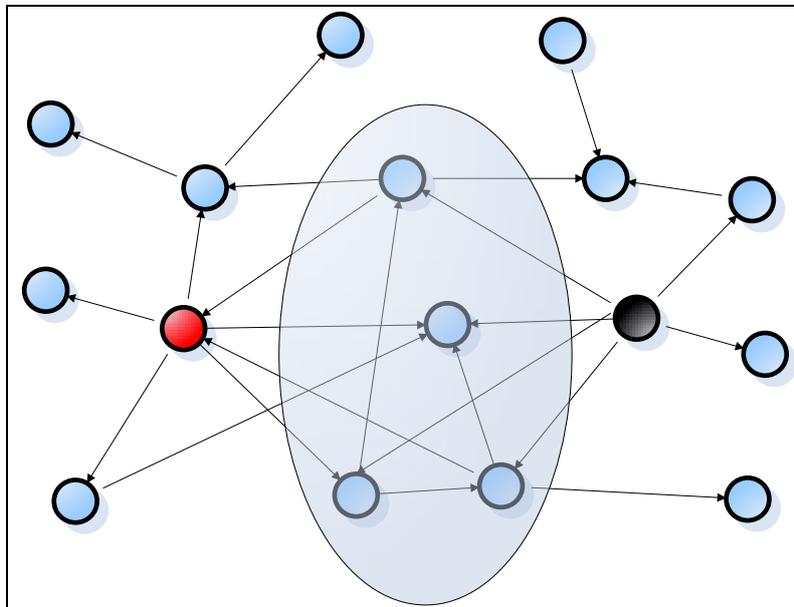


Figura 13 – sub-rede de relacionamentos dos vértices vermelho e preto. Os vértices adjacentes comuns aos dois principais estão circundados.

Na figura acima estão circundados os vértices adjacentes ao vértice vermelho e ou vértice preto ao mesmo tempo. Para esse caso o I_1 tem valor **4**, e o I_2 tem valor **0,5**.

```

--Selecionando o conjunto dos vértices adjacentes ao mesmo tempo dos dois
INSERT @uidsComuns (uid)
    SELECT p1.SourceUidTo FROM dbo.PersonPerson p1
        JOIN dbo.PersonPerson p2 ON p2.SourceUidTo = p1.SourceUidTo
        WHERE p1.SourceUidFrom = @uid1
            and p2.SourceUidFrom = @uid2

--Total de relações do sub grafo formado entre amigos em comum de uid1 e uid2

select @totalReal = count(conjuntoReal.UidFrom) from (
    select pp.SourceUidTo as uidFrom, pp.SourceUidFrom as uidTo
    FROM @uidsComuns pp1 CROSS JOIN @uidsComuns pp2
        JOIN dbo.PersonPerson pp ON pp.SourceUidFrom = pp1.uid
            and pp.SourceUidTo = pp2.uid

UNION

--Aqui faz uma união distinta com o mesmo grupo porém com a direção do
--relacionamento invertida
    select pp.SourceUidFrom as uidFrom, pp.SourceUidTo as uidTo
    FROM @uidsComuns pp1 CROSS JOIN @uidsComuns pp2
        JOIN dbo.PersonPerson pp ON pp.SourceUidFrom = pp1.uid
            and pp.SourceUidTo = pp2.uid

) conjuntoReal

--Total possível de relações nessas combinações
    select @totalPossiveis = COUNT(SourceUidTo) FROM (
        Select p1.SourceUidTo FROM dbo.PersonPerson p1
            JOIN dbo.PersonPerson p2 ON p2.SourceUidTo = p1.SourceUidTo
            WHERE p1.SourceUidFrom = @uid1
                and p2.SourceUidFrom = @uid2

        ) a
    Set @totalPossiveis = @totalPossiveis * (@totalPossiveis - 1)

--A densidade final
Set @retorno = CAST(@totalReal AS decimal(15,5)) / CAST(@totalPossiveis AS decimal(15,5))

```

Figura 14 – Implementação do algoritmo do índice I_2 em T-SQL.

A figura acima mostra a implementação do algoritmo do índice 2 em T-SQL.

3. O terceiro índice é uma variação do segundo. Esse índice mede a densidade do conjunto formado pelos vértices adjacentes ao vértice i e ao vértice j . Ao invés de interseção há uma união dos dois conjuntos de adjacências:

$$I_{3ij} = D_{C_i \cup C_j}$$

Fazendo uma analogia com a rede social analisada, esse índice mede a coesão entre o “grande” grupo formado pelos amigos de i e amigos de j . Um bom índice do tipo 2 não necessariamente acarretará num bom índice do tipo 3. Em outras palavras, uma

pessoa que pertence ao meu ambiente de trabalho, por exemplo, não necessariamente possui amigos que se relacionam de alguma forma com os meus amigos em geral. Neste exemplo, o I_2 possui um bom valor, pois se refere à densidade dos relacionamentos no ambiente de trabalho. Já o I_2 terá um valor pequeno, pois pode não existir relacionamentos entre os dois conjuntos de adjacências. A diferença para o índice 2 é que este procura também por relacionamentos com três passos de distância entre os dois vértices principais. Em outras palavras, quão os amigos do vértice vermelho conhecem os amigos do vértice preto.

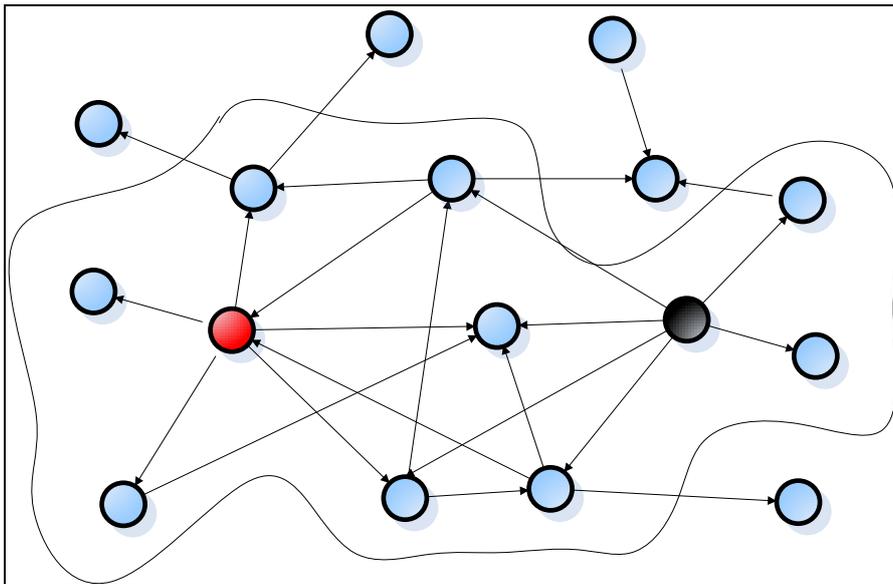


Figura 14 – Os vértices circundados, menos os vértices vermelho e preto, compõem o conjunto analisado no índice 3.

A figura acima mostra o conjunto união dos vértices adjacentes aos vértices vermelho e preto. Para esse caso o valor do I_3 é igual a **0,138**.

```

--total de relações do sub grafo formado entre amigos de uid1 e amigos de uid2
select @totalReal = count(SourceUidFrom) FROM(
  SELECT pp.SourceUidFrom, pp.SourceUidTo FROM dbo.PersonPerson pp1
  CROSS JOIN dbo.PersonPerson pp2
  JOIN dbo.PersonPerson pp ON pp1.SourceUidTo = pp.SourceUidFrom
                           and pp2.SourceUidTo = pp.SourceUidTo
  WHERE pp1.SourceUidFrom = @uid1
        and pp2.SourceUidFrom = @uid2
  UNION --0 inverso
  SELECT pp.SourceUidTo, pp.SourceUidFrom FROM dbo.PersonPerson pp1
  CROSS JOIN dbo.PersonPerson pp2
  JOIN dbo.PersonPerson pp ON pp1.SourceUidTo = pp.SourceUidFrom
                           and pp2.SourceUidTo = pp.SourceUidTo
  WHERE pp1.SourceUidFrom = @uid2
        and pp2.SourceUidFrom = @uid1
) a
  
```

```

--Total possíveis de relações nessas combinações
select @totalPossiveis = COUNT(SourceUidTo) FROM (
    SELECT SourceUidTo FROM dbo.PersonPerson WHERE SourceUidFrom = @uid1
    UNION
    SELECT SourceUidTo FROM dbo.PersonPerson WHERE SourceUidFrom = @uid2
) a

Set @totalPossiveis = @totalPossiveis * (@totalPossiveis - 1)
Set @retorno = CAST(@totalReal AS decimal(15,5)) /
                CAST(@totalPossiveis AS decimal(15,5))

```

Figura 15 – Implementação do algoritmo do índice I_3 em T-SQL.

2.3.2.2. Ponderando os índices

A etapa de ponderação é o momento em que os vários índices, nesse caso são 3, são misturados para juntar um só. Esse valor é utilizado pela ordenação final do conjunto resultante do mecanismo de sugestão. A alternativa mais simples para misturar os índices é utilizando uma média ponderada. Dessa forma, o problema final se reduziu a escolha do peso para cada índice.

A calibragem dos pesos deve ser feita para otimizar o resultado. Otimizar deve ser entendido por classificar os usuários mais importantes no começo da lista. A importância de um usuário da rede social depende do contexto da consulta. O contexto muda dependendo do usuário que está buscando as sugestões. Para isso a função de otimização precisa levar em consideração a estrutura de relacionamentos desse usuário.

Para criação da função de otimização para escolha dos pesos foi feita uma mudança na etapa de filtragem. A etapa de filtragem agora precisa incluir também os vértices relacionados diretamente com o vértice central do processo de sugestão. Pode parecer estranho incluir numa lista de sugestão os vértices que já estão relacionados ao vértice buscador. O propósito é exatamente esse, classificá-los também. A função de otimização utilizará a classificação desses vértices como medida de acerto. Como a ordenação define posições para cada vértice, a média das posições desses vértices, já relacionados ao central, é a medida de otimização. Quanto menor for esse valor, melhor está sendo a ponderação dos índices.

Um mecanismo de auto-ajuste se mostra muito factível para o propósito do mecanismo de sugestão. Esse mecanismo escolherá índices diferentes para usuários

diferentes. Cada usuário tem seu padrão de relacionamentos, e preservar esses padrões é a melhor alternativa para uma sugestão mais eficaz. Um mecanismo de auto-ajuste precisa ser capaz de calibrar os pesos baseando-se na função de otimização. A confecção de tal mecanismo fica como trabalho futuro, uma vez que o escopo desse projeto já suportou muitos tópicos importantes.

```

--Criando tabela que marca as posições
create table #tableOrdenacao(posicao int identity(1,1), uid numeric(20))
INSERT #tableOrdenacao (uid)
    SELECT p.uid FROM person p
        WHERE p.finalScore is not null
            ORDER BY p.finalScore DESC
--Calculando a media da posição dos vértices já relacionados ao vértice @uidSource
SELECT @media = avg(cast(tb.posicao as decimal(10,5))) FROM #tableOrdenacao tb
    JOIN dbo.PersonPerson pp ON pp.SourceUidTo = tb.uid
    WHERE pp.SourceUidFrom = @uidSource

```

Figura 16 – Implementação do algoritmo da função de otimização em T-SQL.

Após a calibragem dos pesos, seja manual ou automática, a lista de usuários a serem sugeridos pode ser gerada a partir da lista dos usuários que foram ponderados removendo aqueles que já possuem relação com o usuário central.

Nos experimentos com os dados coletados e armazenados no banco de dados, a calibragem dos pesos foi feita de forma manual e baseada na função de otimização. Os resultados obtidos foram:

Sub-rede	Peso do I_1	Peso do I_2	Peso do I_3	Qtd de vértices adjacentes ao vértice central	Função de otimização
Brasileira	10	3	180	461	513.60
Indiana	-38	-11	100	535	272.65

Tabela 3 – Resultados do mecanismo de sugestão para as sub-redes brasileiras e indianas.

A tabela acima mostra os resultados obtidos aplicando o mecanismo de sugestão sobre as duas sub-redes armazenadas. Os pesos foram ajustados manualmente. As variações foram curtas até alcançar valores onde a função de otimização atinge o valor melhor, ou seja, o valor menor. Para a sub-rede brasileira o melhor resultado para a função foi 513, isso significa que com os pesos escolhidos, os usuários adjacentes ao usuário central da análise, eles tiveram um posicionamento médio de 513. Já para a sub-

rede indiana esse valor foi melhor, 272. Essa função tem limite mínimo como sendo a metade da quantidade de usuários adjacentes ao usuário central, se esse valor for atingido então todos os usuários já relacionados foram otimamente classificados. O valor da sub-rede indiana tão próximo desse limite mínimo, nesse caso é 267, pode ser sinal de que a conta do usuário central é falsa ou de alguém realmente fanático. São apenas suposições necessitando de comprovações para o fenômeno.

Para que uma avaliação de eficiência que meça a taxa de acerto seja aplicada é necessária posterior validação manual do resultado com o dono da conta do usuário que serviu de elemento central para o processo de sugestão. O usuário central da sub-rede brasileira sou eu! E uma avaliação empírica resultou numa taxa por volta de 75% de acerto nos 20 primeiros. Até os 100 primeiros existiam resultados distribuídos de maneira uniforme. Isso mostra que os resultados estão próximos do topo, porém necessitam de melhorias na precisão para filtrar os erros.

Existem várias possibilidades para melhorias na precisão. Criação de mais outros índices, pré-processamento dos índices para melhorar a normalização e uniformidade das variações, utilização de combinações de pesos que levem a um mínimo global, dentre outros.

Os valores negativos para os pesos dos índices na sub-rede indiana podem ser sinal de alguns padrões de relacionamentos. Que podem ter alguma relação com fatos sociais ou não. A grande variação de escala para os pesos em ambas as sub-redes é esperado, uma vez que numa média ponderada o que importa é a proporção entre os pesos e não o valor de cada um independentemente. Outra razão é que o valor do índice 3 tem, em média, valores centesimais e que não variam de forma uniforme.

3. CONCLUSÕES E TRABALHOS FUTUROS

Este projeto foi concluído com êxito tendo em vista o tamanho do problema inicial. Os resultados foram fantásticos alcançando o propósito inicial do projeto. Muitas dificuldades aconteceram durante o desenvolvimento dele. Por necessitar de grandes recursos computacionais e depender inteiramente da Web, como interface de acesso aos dados, ele se mostrou complicado algumas vezes.

A oportunidade de fazer um estudo aprofundado na área de redes complexas mostrou que muitos problemas podem ser resolvidos quando modelados em forma de uma rede.

A utilização da rede social Orkut mostrou que as possibilidades são inúmeras para os sistemas que se enquadram no modelo social. Diferente de outros tipos de redes, as redes sociais constituem grandes repositórios de informações pessoais. Essas informações podem de várias maneiras, serem utilizadas com fins mais interessantes. Pode servir como fonte de estudos estatísticos populacionais, pode servir como fonte de estudos sociológicos e culturais, além das várias possibilidades de aplicações comerciais. A mais simples, e com certeza já aplicada, é a distribuição de publicidade direcionada a um grupo de usuários que compõe alguma sub-rede.

O desenvolvimento deste projeto ainda está em fase inicial necessitando de inúmeras melhorias e adaptações. Podemos citar algumas:

- Mudança no processo de extração dos dados. A utilização da interface Web para acesso às páginas do Orkut acarreta grande custo computacional, além de muito tempo. Com a nova plataforma para redes sociais e de código aberto lançada pelo Google, Open Social, será possível acesso direto aos dados da estrutura das redes sociais [13].

- Aperfeiçoamento e criação de mais índices que influenciam numa melhor ponderação trazendo mais qualidade aos resultados.
- Construção do mecanismo de auto-ajuste dos pesos dos índices. A utilização de alguma técnica de aprendizagem se faz necessária uma vez que a ponderação dos pesos se mostrou mais interessante do que parecia inicialmente.
- Executar maior número de testes em sub regiões diferentes da rede social, variando de forma organizada.
- Implementação do mecanismo de sugestão dentro de uma rede social existente como o Orkut. Deixando de ser um projeto experimental para ser aplicado e oferecido como mais uma ferramenta da rede social.
- Utilizar o serviço implementado como fonte de *feedback* para melhorar o a aprendizagem do sistema.

BIBLIOGRAFIA

[1] Melanie Mitchell, *Complex Systems: Network Thinking*, Artif. Intell, 170(18), pp. 1194-1212, 2006.

[2] Berkowitz, S. D., *An Introduction to Structural Analysis: The Network Approach to Social Research*. Toronto: Butterworth, 1982.

[3] *Social Network Service*. Acessado em: 20, janeiro, 2008, em: http://en.wikipedia.org/wiki/Social_network_service

[4] Starr Roxanne Hiltz, Murry Turoff. *The Network Nation: Human Communication via Computer*, Addison-Weseley, 1978.

[5] M. E. J. Newman. *Random graphs as models of networks*. In S. Bornholdt and H. G. Schuster, editors, *Handbook of Graphs and Networks*, pages 147{169, Berlin, 2003. Wiley-VCH.

[6] M. E. J. Newman, A.-L. Barabási, and D. J. Watts. *The Structure and Dynamics of Networks*. Princeton University Press, Princeton, NJ, 2006.

[7] D. Watts, S. Strogatz. *Collective dynamics of small-world networks*. *Nature*, 393:440-442, 1998.

[8] Barabási, Albert-László and Réka Albert, *Emergence of scaling in random networks*, *Science*, 286:509-512, October 15, 1999

[9] Réka Albert, Hawoong Jeong, and Albert-László Barabási. *Diameter of the World Wide Web*. *Nature* 401, 130-131 (1999).

[10] Price, D. J. de S., *Networks of scientific papers*, *Science*, 149, 1965, pp. 510-515; Price, D. J. de S., *A general theory of bibliometric and other cumulative advantage processes*, *J. Amer. Soc. Inform. Sci.*, 27, 1976, pp. 292-306.

[11] L. da F. Costa, F. A. Rodrigues, G. Travieso, P. R. Villas Boas, *Characterization of Complex Networks: A survey of measurement*, Advances in Physics, Volume 56, pages 167 - 242, Issue 1 (2007)

[12] N. Alves Jr., M. P. de Albuquerque e M. P. de Albuquerque, J. T. de Assis, *Topology and Shortest Path Length Evolution of The Internet Autonomous Systems Interconnectivity*, Centro Brasileiro de Pesquisas Físicas - CBPF/MCT

[13] *OpenSocial*. Acessado em: 20, janeiro, 2008, em: <http://code.google.com/apis/opensocial/>.