



## **PROPOSTA DE IDENTIFICAÇÃO DA EVOLUÇÃO TEMPORAL DE USO DE APLICATIVOS SOCIAIS PARA A DEFINIÇÃO DE PADRÕES COMPORTAMENTAIS DOS USUÁRIOS**

**Fábio Paschoal Júnior**<sup>1,2</sup>

fabiopjr@yahoo.com.br

<sup>1</sup>. COPPE – Universidade Federal do Rio de Janeiro

Cidade Universitária – CT – Bloco B – Sala 100 – 21945-970, RJ, Rio de Janeiro, Brasil

<sup>2</sup>. CEFET/RJ – Centro Federal de Educação Tecnológica Celso Suckow da Fonseca

Escola de Informática e Computação – Departamento de Informática (DEPIN) – Av. Maracanã, 229, 20271-110, RJ, Rio de Janeiro, Brasil

**Gabriel Vinicius Silva Ribeiro**

**Leandro Moniz de Aragão Daquer**

gabriel.vinicius92@gmail.com

leandro.daquer@outlook.com

CEFET/RJ – Centro Federal de Educação Tecnológica Celso Suckow da Fonseca

Escola de Informática e Computação – Departamento de Informática (DEPIN) – Av. Maracanã, 229, 20271-110, RJ, Rio de Janeiro, Brasil

**Nelson Francisco Favilla Ebecken**

nelson@ntt.ufrj.br

COPPE – Universidade Federal do Rio de Janeiro

Cidade Universitária – CT – Bloco B – Sala 100 – 21945-970, RJ, Rio de Janeiro, Brasil

**Resumo.** A popularização do uso das redes sociais e dos dispositivos móveis, incluindo os acessórios que possibilitam o monitoramento humano, trouxe um ambiente propício para o desenvolvimento de aplicativos sociais (AS) que trabalham com bio-sensores (GPS, acelerômetros, giroscópios, monitores cardíacos, pulseiras inteligentes e relógios inteligentes). Os AS evoluem rapidamente e possibilitam o compartilhamento cada vez maior de informações relevantes nas redes sociais. Nesse sentido, é possível fazer a análise dessas informações com o objetivo de definir algum padrão comportamental do usuário (PCU) em alguma área específica. Este artigo apresenta um estudo em andamento sobre a evolução temporal de uso dos AS na rede social Facebook, o qual engloba a mineração de dados das

publicações dos AS nos perfis dos usuários desta rede social. O estudo de caso está relacionado à mineração dos dados das publicações dos AS de atividades físicas (fitness) e dos perfis dos usuários, para possibilitar a correlação de hábitos saudáveis e atividades físicas, com o intuito da predição do comportamento saudável do usuário. A ferramenta de extração de dados está sendo desenvolvida na forma de um AS para Facebook, tendo como base as tecnologias RestFB, JSoup e Facebook API Graph. O atrativo para o uso desta ferramenta de extração foi a elaboração de rankings competitivos personalizados pelos usuários, os quais podem ser compartilhados no perfil do usuário, para permitir a propagação de uso desta ferramenta e, conseqüentemente, a extração de uma quantidade significativa de dados respeitando as limitações de privacidade dos usuários do Facebook. Assim, será possível a mineração dos dados para a definição de PCU de hábitos saudáveis, com o uso de tecnologias de BigData e de aprendizado de máquina, a partir do uso dos AS apoiados por bio-sensores.

**Palavras:** Mineração de dados, Redes sociais, Facebook, Aplicativos sociais, Dispositivos móveis

# 1 INTRODUÇÃO

As Redes Sociais Online (RSO) estão cada vez mais inseridas no cotidiano da humanidade. Isso é percebido tanto pelo constante crescimento da quantidade de seus usuários quanto pelo aumento do tempo de utilização pelos mesmos. Devido a isso, as RSO já são ótimos instrumentos para estudos e análises do comportamento humano, principalmente pela influência que seus usuários exercem uns sobre os outros, com o compartilhamento de informações, de ideias e de opiniões (Joinson 2008).

As barreiras de localidade entre os usuários das RSO são amenizadas pela comunicação e pela interação entre os mesmos. Assim, as distâncias entre os usuários são irrelevantes e não impedem a crescente interação e, conseqüentemente, o respectivo volume de dados. Portanto, para recuperar conteúdos desse volume de dados são necessários recursos e técnicas computacionais que contenham grande eficiência (Chau et al. 2007).

Dentre as RSO, o Facebook destaca-se por possuir a maior quantidade de usuários (Cosenza 2014). Com essa grande quantidade de usuários, a sociabilidade entre eles cresce e o compartilhamento de conteúdos pode despertar sentimentos e até mesmo um contágio emocional entre os mesmos (Barbosa et al. 2013, Kramer 2012).

Esse contágio emocional ocorre tanto de forma positiva quanto de forma negativa. A forma positiva ocorre quando o usuário passa a ter um comportamento mais solidário ou participativo em relação ao conteúdo compartilhado (Paschoal Jr e Ebecken 2014). A forma negativa ocorre quando o usuário passa a ter um comportamento que motive alguma intolerância, seja racial, cultural, política ou religiosa (Paschoal Jr et al. 2015), bem como a incitação de atos violentos, ou a exposição de crimes ou situações ilícitas, ou até mesmo o planejamento de atividades suspeitas (Silva 2015).

Como os usuários do Facebook compartilham conteúdos variados e de origens diversas, este trabalho demonstra a evolução do trabalho exposto por Paschoal Jr e Ebecken (2014). Porém, com a redução dos estudos de caso propostos pelos mesmos para um único estudo de caso. Assim, o estudo de caso passa a ser somente o de hábitos saudáveis dos usuários do Facebook, especificamente relativos às atividades físicas dos mesmos, sendo descartado o estudo de caso de gostos musicais (Paschoal Jr e Ebecken 2014). O aplicativo de extração das publicações das atividades físicas dos perfis dos usuários do Facebook, denominado FitRank, está em sua fase final de desenvolvimento (Paschoal Jr et al. 2015).

Porém, mesmo com essa redução para um único estudo de caso, a proposta deste trabalho continua a mesma proposta por Paschoal Jr e Ebecken (2014), ou seja, este trabalho se propõe a abordar os seguintes problemas: i.) Criar uma forma automatizada e disponível publicamente para classificar os usuários conforme as características de seu perfil; ii.) criar uma forma automatizada e disponível publicamente para classificar os usuários conforme a utilização de aplicativos sociais (AS), levando em consideração o conteúdo criado a partir desses aplicativos; iii.) criar mecanismos que analisem a evolução temporal de uso dos AS para estabelecer padrões comportamentais dos usuários (PCU).

Além disso, o objetivo também continua o mesmo proposto por Paschoal Jr e Ebecken (2014), ou seja, o objetivo principal deste trabalho é criar uma arquitetura escalável de mineração de dados do Facebook, de forma a classificar os seus usuários conforme perfis de utilização de AS com interação com o Facebook. Com isso, este trabalho busca: i.) usar a mineração de dados para avaliar o padrão de evolução dos AS que os usuários do Facebook

estão utilizando; ii.) classificar os perfis de usuários; iii.) fazer a análise comportamental com a evolução temporal da propagação do uso dos AS.

## **2 REDES SOCIAIS ONLINE**

Atualmente, as RSO podem ser consideradas como um serviço de Internet para que seus usuários façam interações entre si, compartilhando informações, conteúdos e opiniões em perfis pessoais ou empresariais, podendo ser compostas por perfis públicos ou parcialmente públicos (Benevenuto 2010, boyd e Ellison 2008, Mislove 2009).

Mesmo com a existência de mais de 200 RSO (Wikipedia 2014) e de elas terem alcance mundial, há significativas diferenças etnográficas de comportamento de seus usuários. Nesse sentido, encontramos mudanças tanto nas formas de publicação de conteúdos quanto nas formas de interação e nas formas de se auto expressar, de acordo com a localidade dos usuários e de acordo com os objetivos das RSO (Chapman e Lahav 2008).

Dentre as diversas RSO, o Facebook é a mais utilizada (Cosenza 2014) e é o objeto de estudo deste trabalho.

### **2.1 Facebook**

O Facebook possibilita a criação de perfis para seus usuários, tanto para usuários pessoais, quanto para usuários empresariais. Esses perfis contêm mecanismos de compartilhamento de informações que possibilitam a descrição dos atributos do usuário, como nome, idade, amigos, fotos, locais que frequentou, etc. (boyd e Ellison 2008, Ellison et al. 2007), bem como é possível associar os AS ao perfil do usuário, onde esses AS contribuem com o nível de interação do usuário no Facebook.

Esses mecanismos de compartilhamento são atualizados periodicamente, com o incremento de melhorias e novas funcionalidades para motivar o usuário a ter uma maior interação com o Facebook, bem como de atrair novos usuários para essa rede social (Ellison et al. 2007, Joinson 2008, Kraft et al. 2008, Lampe et al. 2008, Olsen e Kraft 2009).

Os conteúdos dos perfis dos usuários baseiam-se em itens compartilhados, como fotos, vídeos, comentários, hipertextos, declarações, músicas, animações, locais frequentados, eventos que participaram ou participarão, atividades que executaram, atualizações de status e outros conteúdos (Barbosa et al. 2013, Olsen e Kraft 2009).

Os AS também criam e compartilham conteúdos no perfil do usuário no Facebook, conforme o propósito de utilização do AS (Paschoal Jr e Ebecken 2014). Todos os itens compartilhados nos perfis dos usuários são passíveis de interações com os seus amigos, seja como curtida, marcações, comentários, compartilhamentos e até mesmo bate-papo (Barbosa et al. 2013, Olsen e Kraft 2009).

Como o Facebook está constantemente aprimorando os seus mecanismos para aumentar a privacidade de seus usuários, a extração de dados no Facebook tem sido um desafio. Os AS que desejarem extrair informações devem solicitar ao Facebook as permissões de acesso aos tipos de conteúdos que serão extraídos pelos AS, bem como é necessário solicitar a autorização dos usuários para os AS acessarem os dados privados dos mesmos (Cheng et al. 2013, Paschoal Jr et al. 2015).

## **2.2 Aplicativos Sociais com uso em smartphones**

Os smartphones fornecem aos AS uma característica importante de alguns sistemas informatizados, que é a ubiquidade ou a computação ubíqua. Isso significa que os AS, com o uso em smartphones, estão disponíveis a qualquer hora, em qualquer lugar e são utilizados para muitas atividades distintas. Os usuários utilizam os AS de forma prática e simples, sem perceber a complexidade embutida nos AS, sendo transparente aos seus usuários as barreiras computacionais e as complexidades de execução das tarefas (Domingues 2008).

Nesse sentido, os AS usam as conexões dos smartphones para conectarem-se à Internet ou a outros sistemas computacionais, acessando ou disponibilizando informações na Internet ou em sistemas específicos de geo-referenciamento (Paschoal Jr e Ebecken 2014).

Assim, o uso dos AS em smartphones, associados com o uso das RSO, estão tornando verdadeiras ferramentas persuasivas para estimular a mudança do estilo de vida dos usuários dos AS e das RSO (Morris et al. 2011).

Essa persuasão pode ser motivada pelo bio-sensoriamento móvel dos usuários dos AS em smartphones. Isso significa que os AS podem disponibilizar algumas propriedades de seus usuários, relatando as mesmas de forma temporal conforme a utilização dos AS com o passar do tempo. Esse bio-sensoriamento enriquece o uso dos AS e fornecem dados dos detalhes das atividades realizadas pelos usuários nos AS (Paschoal Jr e Ebecken 2014).

Portanto, os sensores embutidos nos smartphones ou os sensores externos que podem ser acoplados aos mesmos, como o GPS, o acelerômetro, o giroscópio, o monitor cardíaco, as pulseiras inteligentes ou os relógios inteligentes, possibilitam aos AS a obtenção das propriedades do usuário durante a realização das atividades com o uso do AS. Neste trabalho, esses sensores serão considerados como bio-sensores (Paschoal Jr e Ebecken 2014).

Consequentemente, o bio-sensoriamento móvel feito pelos AS em smartphones e disponibilizado nas RSO possibilita o estabelecimento de PCU, permitindo aos usuários uma experiência de monitoramento mais detalhada com a disponibilização de informações importantes dos usuários (Paschoal Jr e Ebecken 2014).

## **2.3 Aplicativos Sociais**

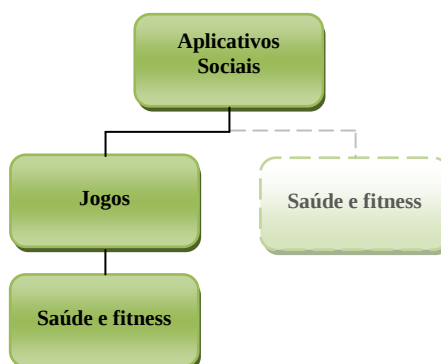
Os AS possibilitam a integração de seus dados com o Facebook e permitem uma maior socialização nesta rede social, através do compartilhamento e da recuperação de informações nos perfis dos usuários do Facebook. Eles podem ser desenvolvidos como páginas internas ou externas ao Facebook, com o uso da Graph API do Facebook (Facebook 2015a).

Esta API possibilita a publicação de conteúdos, histórias e fotos no perfil do usuário do Facebook, bem como consultar as propriedades dos usuários e os conteúdos das publicações disponíveis nos perfis dos usuários (Facebook 2015a).

Uma história é uma publicação especial que permite a associação de ações do usuário a objetos pré-estabelecidos como conteúdos de publicações do Facebook como, por exemplo, o mapa com o geo-referenciamento onde a atividade física foi realizada (Facebook 2015b).

O Facebook faz a categorização dos AS conforme os objetivos e finalidades dos mesmos. Há uma extensa lista de AS disponíveis no Facebook (Facebook 2014) e a Figura 1 simplifica a categoria dos AS do estudo de caso deste trabalho.

Recentemente, o Facebook reclassificou a maioria dos AS de atividades físicas dentro da categoria de Jogos, ilustrado na Figura 1. Na especificação desses AS eles ainda continuam como aplicativos de saúde e fitness, porém não há mais essa categoria independente de Saúde e Fitness na lista de AS do Facebook e sim em uma subcategoria de Jogos (Endomondo 2015, Nike 2015a, RunKeeper 2015, Runtastic 2015a).



**Figura 1 – Categorias simplificadas de aplicativos sociais do Facebook**

Através do uso desses AS e das postagens feitas pelos mesmos nos perfis de seus usuários no Facebook, ocorre o enriquecimento na disponibilização dos dados desses usuários no sentido de ser possível inferir características de seu comportamento, hábitos, preferências e gostos (Paschoal Jr e Ebecken 2014).

Essas postagens compartilhadas pelos AS possuem formatos padronizados, dentro de algumas variações estabelecidas pelo Facebook, segundo a plataforma computacional utilizada no compartilhamento e o tipo de informação disponibilizada pelo AS. Informações como hipertextos, mensagens, atualizações de status ou uma história estruturada (como uma atividade física realizada) são os tipos de informações que podem ser compartilhadas (Facebook 2015c).

Também encontramos postagens compartilhadas como histórias mais elaboradas e personalizadas, seguindo uma estrutura definida, como o ator que faz a publicação, a ação que está sendo relatada pelo ator, o objeto da ação descrevendo como ela foi realizada pelo ator, bem como o AS que fez a publicação. Essas postagens mais elaboradas são feitas pelos AS que utilizam a ferramenta Open Graph do Facebook (Facebook 2015b).

## **2.4 Extração de dados do Facebook**

O Facebook disponibiliza recursos computacionais que permitem o acesso ao conteúdo das informações disponibilizadas por seus usuários em seus perfis. Estes recursos são conhecidos como Application Programming Interfaces (APIs) e definem a forma de comunicação e de troca de dados dos AS com o Facebook (Paschoal Jr e Ebecken 2014).

As APIs do Facebook são atualizadas periodicamente para incorporar as novas funcionalidades que são disponibilizadas nesta rede social. As versões dessas APIs têm suporte por 2 anos e sofrem evoluções em um prazo menor. Assim, a versão inicial v.1.0 foi criada em 2010, que evoluiu para a versão 2.0 em 2014. Em pouco mais de um ano tivemos as versões 2.1, 2.2, 2.3 e 2.4. A atual versão 2.4 será suportada até agosto de 2017 (Facebook 2015d, 2015e).

A principal API do Facebook é a Graph API, a qual possibilita a postagem e a recuperação de informações nos perfis dos usuários do Facebook (Facebook 2015a). Para

isso, os AS são construídos para fazerem a comunicação com o Facebook seguindo as regras de comunicação desta API, bem como mediante as autorizações do Facebook e de seus usuários para acessar as informações. As respostas dessa API podem ser feitas tanto no formato JavaScript Object Notation (JSON), quanto no formato Extensible Markup Language (XML) (Paschoal Jr e Ebecken 2014).

Há uma API alternativa para a extração dos dados do Facebook, a API RestFB, a qual não foi desenvolvida pelo Facebook. Ela facilita esse trabalho de extração, por ser robusta às constantes modificações da Graph API, possibilitando uma maior estabilidade ao desenvolvimento dos AS. Assim, as operações disponibilizadas pela API RestFB não sofrem constantes alterações e essa API absorve as modificações da Graph API. A API RestFB trabalha no ambiente Java e permite aos AS o trabalho em conjunto com a Graph API (Allen 2015).

Os AS de atividades físicas compartilham postagens, nos perfis de seus usuários no Facebook, com um hipertexto para o site do AS. Isso significa que as postagens contêm um hipertexto para um site externo ao Facebook, detalhando a execução da atividade física de seus usuários, com informações complementares ao que foi postado no Facebook. A API JSoup pode ser utilizada para fazer a extração dessas informações externas ao Facebook, já que ela tem a função de extrair informações de páginas da web. A JSoup é uma biblioteca desenvolvida em Java e é de uso livre, de código aberto (Hedley 2015).

## **2.5 Mineração dos dados extraídos do Facebook**

O processo de mineração dos dados extraídos do Facebook e o tratamento desses dados minerados permitem a realização de análises para o estudo dos PCU, além de identificar gostos e interesses dos usuários desta rede social. Para isso, é importante identificar os atributos relevantes, como pessoas, atividades ou localidades (Benevenuto et al. 2012).

Devido ao grande número de usuários do Facebook e do crescente nível de socialização dos mesmos, disponibilizando um volume de dados cada vez maior, a mineração desses dados representa um problema de BigData (Jagadish et al. 2014), consumindo grandes recursos computacionais para o seu processamento. Para processar esse grande volume de dados heterogêneos (BigData), as tecnologias de Hadoop (Noll 2011, The Apache Software Foundation 2015a), MapReduce (Dean e Ghemawat 2008), HBase (The Apache Software Foundation 2015b), Mahout (The Apache Software Foundation 2014a) são as atuais tecnologias que possibilitam uma redução do tempo de processamento, comparado com o processamento em um computador convencional. Assim, é possível escalar o processamento conforme a necessidade e a demanda de processamento dos dados (Paschoal Jr e Ebecken 2014).

## **3 ARQUITETURA**

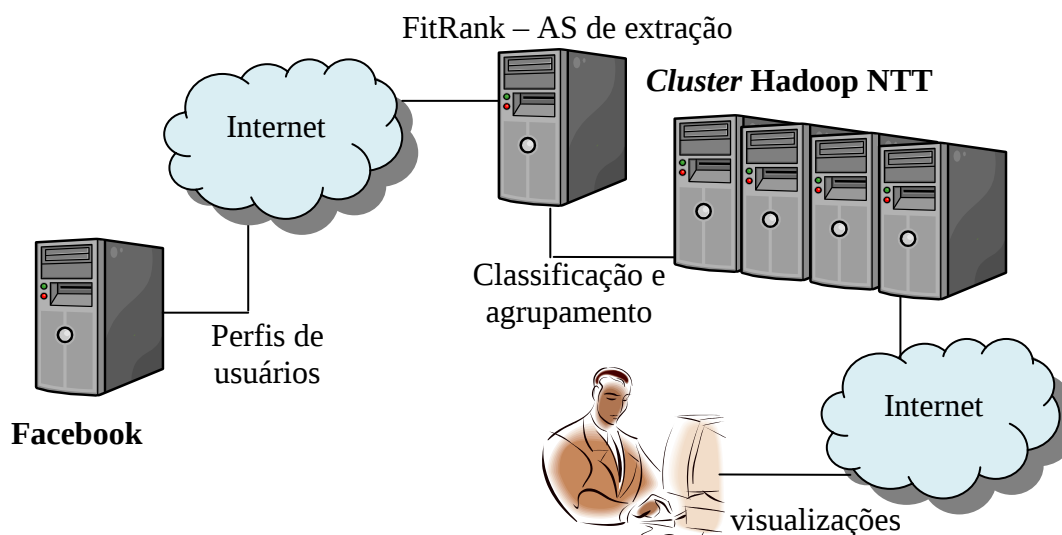
A Figura 2 demonstra a arquitetura computacional deste trabalho. O Facebook disponibiliza o acesso aos perfis de seus usuários para o FitRank, que é o AS de extração de dados, o qual armazenará as informações das atividades físicas coletadas no banco de dados hospedado no Cluster Hadoop NTT (HBase).

Posteriormente, esses dados serão minerados para permitirem as análises dos PCU. Assim, as ferramentas de classificação e de agrupamento (Mahout) executarão as suas

atividades no Cluster Hadoop NTT, com processamento paralelo e distribuído (Hadoop e MapReduce).

A escolha do banco de dados será o HBase, pelo motivo de ter fácil integração com o ambiente Hadoop e de trabalhar com processamento paralelo e distribuído.

Os resultados de processamento dos PCU ficarão armazenados Cluster Hadoop NTT (HBase). Posteriormente, a ferramenta de visualização irá gerar as visualizações com base nesses resultados e de acordo com as solicitações de visualizações feitas pelo usuário.



**Figura 2 – Arquitetura proposta**

### **3.1 FitRank - AS de coleta de dados do Facebook**

O FitRank é um AS que faz a extração de compartilhamentos postados nos perfis de seus usuários no Facebook, relativos às atividades físicas feitas pelos usuários e que foram monitoradas por AS de atividades físicas (Paschoal Jr et al. 2015).

Até a data final de revisão deste artigo, o FitRank está em fase final de desenvolvimento e fornecerá os dados para a composição dos PCU que formam o estudo de caso deste trabalho. Inicialmente, o FitRank capturará as informações geradas por um conjunto de sete AS de atividades físicas (Paschoal Jr et al. 2015), os quais são o Nike+ Running (Nike 2015a, 2015b), o Runtastic Corrida e Caminhada (Runtastic 2015a, 2015b), o Runtastic PRO GPS Correr (Runtastic 2015c), o Runtastic Road Bike Ciclismo (Runtastic 2015d), o Runtastic Road Bike PRO GPS (Runtastic 2015e), o Runtastic Mountain Bike GPS (Runtastic 2015f) e o Runtastic Mountain Bike PRO (Runtastic 2015g), para a validação da metodologia proposta por este trabalho. Posteriormente, ele poderá ser revisto para abranger a inclusão da captura dos dados de outros AS de atividades físicas. Assim, futuramente, será necessário um procedimento de revisão periódica para adicionar a captura dos dados de novos AS de atividade físicas.

A Figura 3 demonstra os tipos de postagens feitas pelos AS de atividades físicas nos perfis dos usuários do Facebook, no formato de postagem de estória, e que poderão ser passíveis de extração de seus dados pelo FitRank, caso esses usuários se tornem usuários do FitRank.



Figura 3 – Postagens de AS de atividades físicas

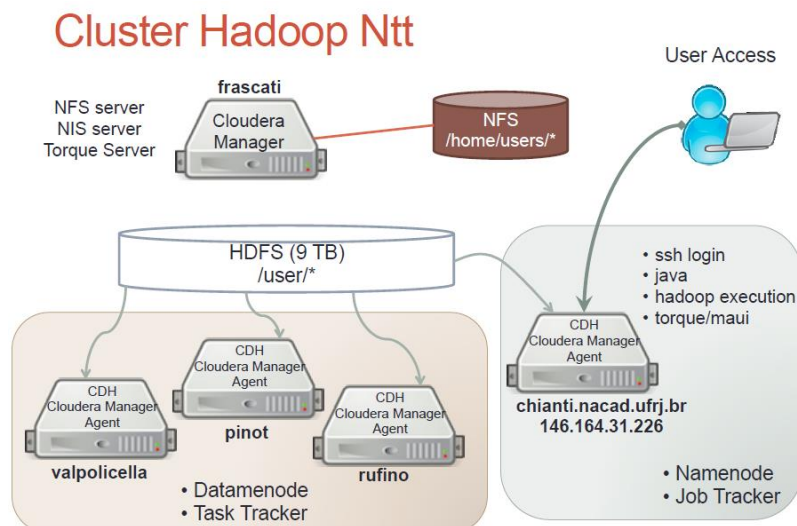
O FitRank trabalhará com o processamento paralelo e distribuído para otimizar a extração de dados do Facebook, de forma a armazenar os seus dados em tabelas do HBase, para permitir as análises dos PCU através da análise da evolução temporal da propagação dos usos das aplicações sociais.

### 3.2 Processamento Paralelo e Distribuído

Como será esperado o processamento de um grande volume de dados heterogêneos (BigData) oriundos da extração de dados pelo FitRank, será necessário utilizar os recursos de processamento de alto desempenho, ou seja, o processamento paralelo e distribuído de um cluster computacional.

O cluster computacional do Núcleo de Transferência de Tecnologia (NTT) da Universidade Federal do Rio de Janeiro (UFRJ) será utilizado para o processamento paralelo e distribuído, o qual disponibiliza o Hadoop, o MapReduce, o HBase e o Mahout. Este cluster computacional é ilustrado na Figura 4.

As ferramentas que serão executadas neste cluster computacional serão desenvolvidas na linguagem Java, a qual é suportada pelo mesmo.



#### **Figura 4 – Cluster Hadoop NTT**

Para garantir a maximização da escalabilidade e do paralelismo, as ferramentas de extração, de classificação e de agrupamento serão executadas seguindo o paradigma de MapReduce.

A classificação, o agrupamento e a sumarização dos dados coletados serão realizados pelos algoritmos de aprendizado de máquina do Mahout, onde serão selecionados os algoritmos que trabalhem com o paralelismo de MapReduce.

A classificação dos dados será feita pelo algoritmo “Naive Bayes” do Mahout (The Apache Software Foundation 2014b). O agrupamento e a sumarização dos dados serão feitos pelo algoritmo “K-Means Clustering” do Mahout (The Apache Software Foundation 2014c). Esses algoritmos serão escolhidos por oferecerem suporte ao paradigma de MapReduce.

### **3.3 Estudo de caso**

O estudo de caso deste trabalho será feito com a mineração dos dados descritos na seção “2.5 Mineração dos dados extraídos do Facebook”, usando o aplicativo descrito na seção “3.1 FitRank - AS de coleta de dados do Facebook” deste trabalho. Porém, esse artigo relata o trabalho em andamento e os resultados surgirão após a utilização do FitRank pelos usuários do Facebook, o que acontecerá em breve.

Os compartilhamentos feitos por quaisquer desses AS de atividades físicas no perfil do usuário, já descritos neste trabalho, serão alvo da extração e da mineração de dados.

Será feita uma análise da evolução temporal de uso desses AS minerados, tanto pelos usuários quanto por seus amigos, para fazer a correlação da influência que um usuário teve com o outro na utilização desses AS.

Neste estudo de caso de hábitos saudáveis, a proposta será analisar os hábitos esportivos dos usuários com base nas informações do perfil do usuário, na seção de atividades físicas, extraíndo essas informações. Além disso, serão analisadas as postagens compartilhadas no mural de publicações do usuário, feitas particularmente pelos AS relativos à prática de atividades físicas.

Serão armazenadas as informações do usuário, as informações da atividade física postada (no perfil ou no mural de publicações), as informações temporais da postagem, as informações de geo-referenciamento da execução da atividade física, as informações complementares do monitoramento da atividade física que estarão disponibilizadas externamente ao Facebook e as informações do aplicativo que fizer a postagem.

As informações de geo-referenciamento permitirão análises das localidades de execução das atividades físicas e possibilitarão a mineração de dados para a descoberta de características como, por exemplo, a distribuição regional de execução das atividades físicas, a frequência de locais para a prática dessas atividades, os horários de pico, a comunidade de usuários de uma determinada localidade, dentre outras descobertas.

Posteriormente, serão executadas as ferramentas de classificação e de agrupamento, já descritas na seção “3.2 Processamento Paralelo e Distribuído” deste trabalho, onde ocorrerá a predição do comportamento saudável dos usuários (pouco saudável, saudável ou muito saudável). Com essas informações, poderá ser feita uma recomendação de como está o comportamento saudável do usuário e como estará futuramente, caso ele mantenha a frequência atual de atividades físicas. Também poderá ser recomendado um novo

comportamento para melhorar o nível saudável do usuário, indicando o quanto ele deve melhorar a sua frequência de atividades (quantidade de atividades, desempenho nas atividades ou duração das atividades).

Essas recomendações serão baseadas nas recomendações de estilo de vida saudável associados aos cuidados com a saúde em relação ao sedentarismo, onde é recomendada a prática regular de atividades físicas (World Health Organization 2002). Nesse sentido, a falta da prática regular de atividades físicas é um dos dez principais fatores de risco de morte, bem como é um fator de risco para doenças cardiovasculares, câncer e diabetes. Atualmente, aproximadamente 25% da população mundial adulta e 80% da população mundial adolescente são consideradas sedentárias (World Health Organization 2015).

A Organização Mundial de Saúde (OMS) (World Health Organization 2010, 2015) define que as pessoas sedentárias têm de 20% a 30% de aumento do risco de morte em relação às pessoas que fazem atividades físicas. Já as pessoas que fazem atividades físicas regularmente têm diversos benefícios em sua saúde, como:

- melhora da situação cardiorrespiratória;
- fortalecimento dos músculos;
- melhora da saúde óssea;
- redução de hipertensão;
- redução de doenças coronárias;
- redução de derrames;
- redução de diabetes;
- redução do câncer de mama;
- redução do câncer de cólon;
- redução de depressão;
- redução do risco de quedas;
- redução do risco de fratura de vértebras;
- melhora do controle do peso;
- melhora do equilíbrio de energia do organismo.

A Tabela 1 representa as recomendações da prática de atividades físicas feitas pela OMS (World Health Organization 2010, 2015), as quais darão base à predição do comportamento saudável dos usuários (pouco saudável, saudável ou muito saudável).

**Tabela 1. Recomendações da OMS sobre a prática de atividades físicas**

| <b>Faixa Etária</b> | <b>Recomendações da OMS</b>                                                                                                                                                                                                    |
|---------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| De 5 a 17 anos      | <ul style="list-style-type: none"><li>• Ao menos 60 minutos de atividades físicas moderadas ou intensas por dia.</li><li>• Atividades físicas com mais de 60 minutos por dia fornecem benefícios adicionais à saúde.</li></ul> |

|                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                     | <ul style="list-style-type: none"><li>• Ao menos 3 vezes por semana atividades físicas que fortaleçam músculos e ossos.</li></ul>                                                                                                                                                                                                                                                                                                                                                                                          |
| De 18 a 64 anos     | <ul style="list-style-type: none"><li>• Ao menos 150 minutos de atividades físicas moderadas ou 75 minutos de atividades físicas intensas por semana.</li><li>• 300 minutos de atividades físicas moderadas por semana fornecem benefícios adicionais à saúde.</li><li>• Ao menos 2 vezes por semana atividades físicas que fortaleçam músculos ou que envolvam uma quantidade significativa de músculos.</li></ul>                                                                                                        |
| A partir de 65 anos | <ul style="list-style-type: none"><li>• Ao menos 150 minutos de atividades físicas moderadas ou 75 minutos de atividades físicas intensas por semana.</li><li>• 300 minutos de atividades físicas moderadas por semana fornecem benefícios adicionais à saúde.</li><li>• Ao menos 3 vezes por semana atividades físicas para melhorar o equilíbrio e prevenir quedas.</li><li>• Ao menos 2 vezes por semana atividades físicas que fortaleçam músculos ou que envolvam uma quantidade significativa de músculos.</li></ul> |

Assim, a predição para o comportamento pouco saudável dar-se-á quando o usuário ficar aquém da quantidade mínima de minutos semanais, da frequência e da intensidade das atividades físicas, conforme a sua faixa etária. A predição para o comportamento saudável dar-se-á quando o usuário atingir a quantidade mínima de minutos semanais, da frequência e da intensidade das atividades físicas, conforme a sua faixa etária. A predição para o comportamento muito saudável dar-se-á quando o usuário atingir da quantidade mínima de minutos semanais para ter benefícios adicionais à saúde, da frequência e da intensidade das atividades físicas, conforme a sua faixa etária.

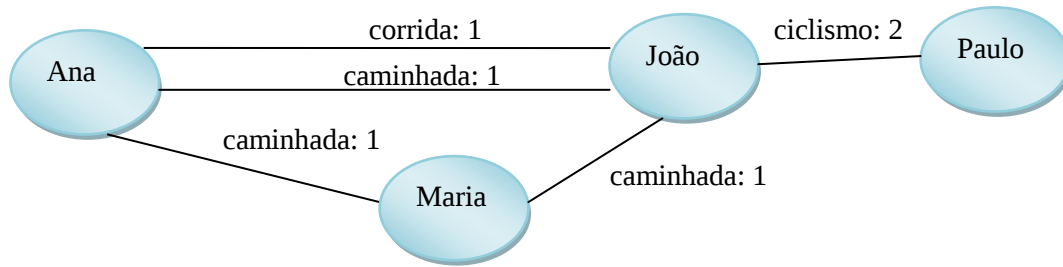
### **3.4 Visualizações**

As visualizações dos dados são importantes para a compreensão dos PCU e, nesse sentido, são ótimas ferramentas para a compreensão dos resultados obtidos. Uma forma de simplificar a visualização, no contexto de RSO, é utilizar a técnica de seleção de partes do grafo, trabalhando com subgrafos (Freitas et al. 2008).

Assim, a visualização pode conter variações e contextos diferentes de apresentação, conforme a necessidade do usuário final. Os subgrafos serão construídos com base na classificação dos dados dos usuários do Facebook, conforme as informações que forem possíveis de serem obtidas, agrupadas e classificadas.

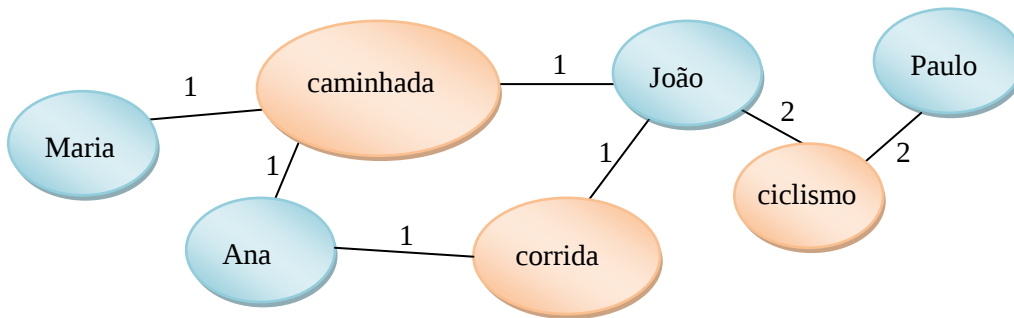
Como exemplo de aplicação desta metodologia, o primeiro contexto de hábitos saudáveis poderá ser construído com vértices representando os usuários do Facebook e as arestas indicando os resultados de similaridade da quantidade de atividades físicas semanais entre eles, identificando inclusive a atividade física. Assim, cada vértice pode ter uma um mais

arestas de interligação, onde a aresta poderá representar a quantidade semanal de atividades físicas que ambos os vértices compartilham. A Figura 5 ilustra a abordagem desse contexto.



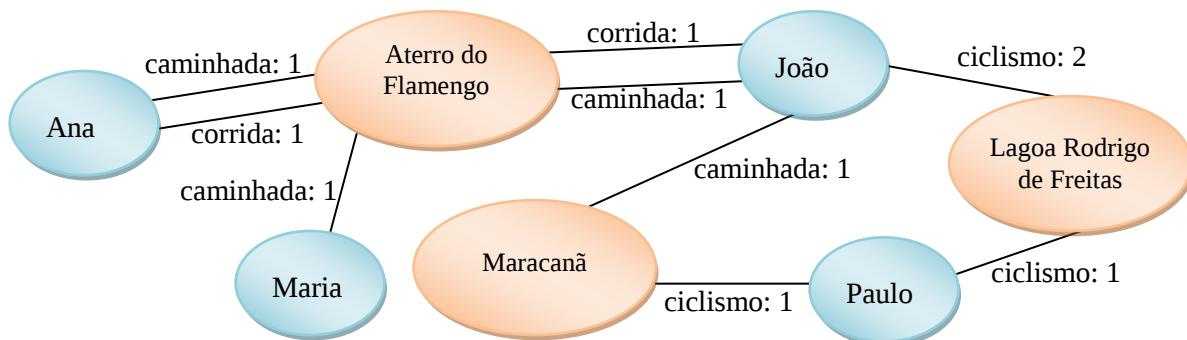
**Figura 5 – Primeiro contexto de visualização de usuários x hábitos saudáveis**

O segundo contexto de hábitos saudáveis poderá ser construído com vértices representando os usuários do Facebook e as atividades físicas, onde as arestas indicam a ligação da atividade física com os usuários. Nesse contexto, há uma rotulagem das arestas e essa rotulagem indica a quantidade semanal que a atividade física foi realizada pelo usuário. A Figura 6 ilustra a abordagem desse contexto.



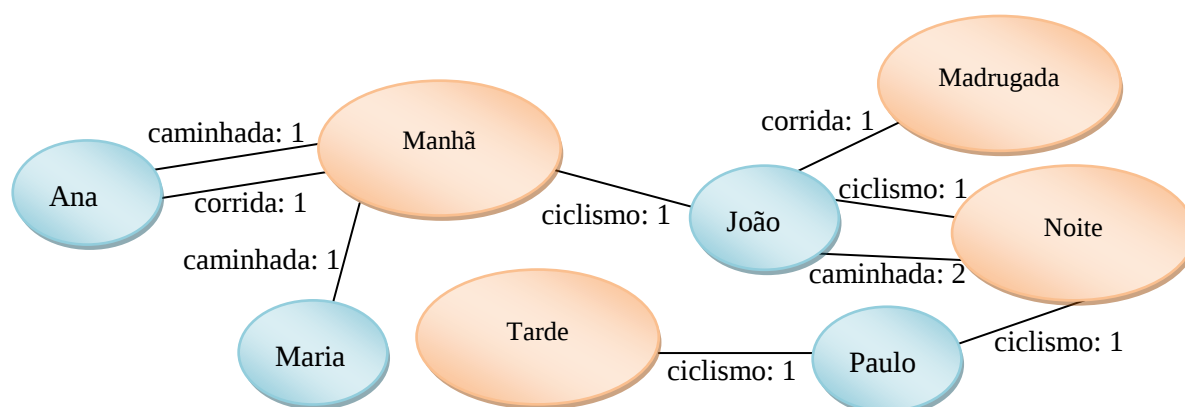
**Figura 6 – Segundo contexto de visualização de usuários x hábitos saudáveis**

O terceiro contexto de hábitos saudáveis poderá ser construído com vértices representando os usuários do Facebook e as localidades onde foram executadas as atividades físicas, onde as arestas indicam a ligação das atividades físicas dos usuários com as localidades. Nesse contexto, há uma rotulagem das arestas e essa rotulagem indica a quantidade semanal que a atividade física foi realizada pelo usuário naquela localidade. O usuário poderá ter mais de uma aresta ligando à localidade, caso ele tenha feito mais de uma modalidade de atividade física naquela localidade. A Figura 7 ilustra a abordagem desse contexto.



**Figura 7 – Terceiro contexto de visualização de usuários x hábitos saudáveis**

O quarto contexto de hábitos saudáveis poderá ser construído com vértices representando os usuários do Facebook e os períodos de dia em que foram executadas as atividades físicas, onde as arestas indicam a ligação das atividades físicas dos usuários com o período de dia. Nesse contexto, há uma rotulagem das arestas e essa rotulagem indica a quantidade semanal que a atividade física foi realizada pelo usuário naquele período de dia. O usuário poderá ter mais de uma aresta ligando ao período do dia, caso ele tenha feito mais de uma modalidade de atividade física no mesmo período de dia. A Figura 8 ilustra a abordagem desse contexto.



**Figura 8 – Quarto contexto de visualização de usuários x hábitos saudáveis**

Esses quatro contextos exemplificados não esgotam as possibilidades de contextos possíveis que são passíveis de visualização. Neles, é facilitada a predição sobre se o usuário irá ter um comportamento saudável pelo fato de que ele estará ligado a uma propriedade que identifica a realização das atividades físicas do mesmo, seja na forma de uma aresta, seja na forma de um vértice.

### **3.5 Avaliação da metodologia**

A avaliação do trabalho será feita por meio da verificação dos acertos de predição do estudo de caso, conforme a classificação de hábitos saudáveis e, neste caso, este trabalho fará a predição de avaliação do comportamento do usuário e o quanto ele é estimulado a ter um comportamento saudável com base nas postagens feitas por amigos a partir de AS de atividades físicas, ou seja, avaliará se ele estará com um comportamento pouco saudável, saudável ou muito saudável. Será feito o levantamento estatístico dos acertos e dos erros da predição.

Como a metodologia deste trabalho faz a recomendação de como estará o comportamento saudável do usuário futuramente, de acordo com a frequência de atividades, a quantidade de atividades, o desempenho das atividades e a duração das atividades, será considerado um acerto da predição se o usuário atingir a recomendação no período previsto e será considerado um erro da predição se o usuário não atingir a recomendação no período previsto.

Além disso, será feita a medida de influência que os usuários provocaram uns nos outros, em relação a ter um comportamento mais saudável, correlacionando os dados de utilização do FitRank com as mudanças de utilização dos AS de atividades físicas, bem como correlacionando a evolução de desempenho da realização dessas atividades físicas motivadas pela competição social que será motivada por esse trabalho.

## 4 CONSIDERAÇÕES PARCIAIS

Este trabalho ainda está em fase inicial de desenvolvimento e, nesse sentido, espera-se que o AS FitRank provoque um efeito viral de utilização do mesmo, motivado pela natureza humana de competitividade e pela necessidade de as pessoas passarem a ter comportamentos mais saudáveis.

Dessa forma, poderá haver uma grande quantidade de dados heterogêneos para ser processada, o que será um problema de BigData. Assim, poderá ser construída pelo FitRank uma ótima base de dados para a mineração dos dados para compor os PCU de atividades físicas.

Portanto, a contribuição deste trabalho possibilitará o estudo da evolução temporal de uso dos AS com o objetivo de estabelecer os PCU de atividades físicas. Ainda assim, com ajustes nas ferramentas que serão desenvolvidas neste trabalho, será possível trabalhar com novos estudos de caso, ampliando a utilização da metodologia proposta por este trabalho.

Com o paralelismo proposto na arquitetura deste trabalho é esperada uma melhora considerável na mineração dos dados das RSO, tendo em vista que o volume de dados heterogêneos da RSO crescem rapidamente e isso é um problema de BigData.

## AGRADECIMENTOS

O autor deste trabalho agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo fomento de bolsa de doutorado ao mesmo.

## REFERÊNCIAS

- Allen, M., (2015). RestFB. Disponível em: <http://restfb.com/>. Acesso em: 14 jul 2014.
- Barbosa, G. A. R., Santos, G. E. dos, Pereira, V. M. de O., (2013), "Caracterização qualitativa da sociabilidade no Facebook", *IHC '13 Proceedings of the 12th Brazilian Symposium on Human Factors in Computing Systems*, p. 72–81.
- Benevenuto, F., (2010), "Redes Sociais Online: Técnicas de Coleta, Abordagens de Medição e Desafios Futuros", *Short course (book chapter) in SBSC'10, Webmedia'10, IHC'10 and SBBD'10*
- Benevenuto, F., Almeida, J. M., Silva, A. S., (2012), "Coleta e Análise de Grandes Bases de Dados de Redes Sociais Online", *Book Chapter in Jornada de Atualizações em Informática (JAI), Congresso da Sociedade Brasileira de Computação (CSBC)*
- boyd, danah m., Ellison, N. B., (2008), "Social network sites: Definition, history, and scholarship", *Journal of Computer-Mediated Communication*, v. 13, p. 210–230.
- Chapman, C. N., Lahav, M., (2008), "International ethnographic observation of social networking sites", *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, p. 3123–3128.
- Chau, D. H., Pandit, S., Wang, S., Faloutsos, C., (2007), "Parallel crawling for online social networks", *World Wide Web Conference (WWW)*, p. 1283–1284.
- Cheng, Y., Park, J., Sandhu, R., (2013), "Preserving user privacy from third-party applications in online social networks", *Proceedings of the 22nd international conference on World Wide Web companion (WWW' 13)*

- Cosenza, V., (2014). World Map of Social Networks. Disponível em: <http://vincos.it/world-map-of-social-networks/>. Acesso em: 14 jul 2015.
- Dean, J., Ghemawat, S., (2008), "MapReduce: simplified data processing on large clusters", *Communications of the ACM*, v. 51, n. 1, p. 107–113.
- Domingues, F. L., (2008), *Computação ubíqua* Disponível em: <http://www.hardware.com.br/artigos/computacao-ubiqua/>.
- Ellison, N. B., Steinfield, C., Rocha, I., (2007), "The benefits of Facebook “friends”: Exploring the relationship between college students’ use of online social networks and social capital", *Journal of Computer-Mediated Communication*, v. 12
- Endomondo, (2015). Endomondo Sports Tracker. Disponível em: <https://www.facebook.com/games/endoapp/?fbs=133>. Acesso em: 28 ago 2015.
- Facebook, (2014). Central de Aplicativos. Disponível em: <https://www.facebook.com/appcenter/>. Acesso em: 1 jul 2014.
- Facebook, (2015a). The Graph API. *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/graph-api>. Acesso em: 14 jul 2015.
- Facebook, (2015b). Creating Custom Open Graph Stories. *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/sharing/opengraph/custom>. Acesso em: 14 jul 2015.
- Facebook, (2015c). Sharing. *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/sharing>. Acesso em: 14 jul 2015.
- Facebook, (2015d). Facebook Platform Versioning. *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/apps/versions>. Acesso em: 14 jul 2015.
- Facebook, (2015e). Facebook Platform Changelog. *Facebook Developers*. Disponível em: <https://developers.facebook.com/docs/apps/changelog>. Acesso em: 14 jul 2015.
- Freitas, C. M. D. S., Nedel, L. P., Galante, R., Lamb, L. C., Spritzer, A. S., Fujii, S., Oliveira, J. P. M. de, Araújo, R. M., Moro, M. M., (2008), "Extração de Conhecimento e Análise Visual de Redes Sociais", *Anais do XXVIII Congresso da SBC – SEMISH – Seminário Integrado de Software e Hardware*, p. 106–120.
- Hedley, J., (2015). jsoup Java HTML Parser, with best of DOM, CSS, and jquery. Disponível em: <http://jsoup.org/>. Acesso em: 20 ago 2015.
- Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., Shahabi, C., (2014), "Big data and its technical challenges", *Communications of the ACM*, v. 57, n. 7, p. 86–94.
- Joinson, A. N., (2008), "Looking at, looking up or keeping up with people?: motives and use of facebook", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 1027–1036.
- Kraft, P., Drozd, F., Olsen, E., (2008), "Digital Therapy: Addressing Willpower as Part of the Cognitive-Affective Processing System in the Service of Habit Change", *PERSUASIVE 2008*, p. 177–188.
- Kramer, A. D. I., (2012), "The spread of emotion via facebook", *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, p. 767–770.

Lampe, C., Ellison, N. B., Steinfield, C., (2008), "Changes in use and perception of facebook", *Proc. Of the 2008 ACM conference on Computer supported cooperative work (CSCW '08)*, p. 721–730.

Mislove, A. E., (2009), *Online Social Networks: Measurement, Analysis, and Applications to Distributed Information Systems*, Rice University Disponível em: <http://www.ccs.neu.edu/home/amislove/publications/SocialNetworks-Thesis.pdf>.

Morris, M. E., Consolvo, S., Munson, S., Patrick, K., Tsai, J., Kramer, A. D. I., (2011), "Facebook for health: opportunities and challenges for driving behavior change", *Proceeding CHI EA '11 CHI '11 Extended Abstracts on Human Factors in Computing Systems* (nov.), p. 443–446.

Nike, (2015a). Nike+ Running - get more from your run. Disponível em: <https://www.facebook.com/games/nikeapp/?fbs=133>. Acesso em: 28 ago 2015.

Nike, (2015b). Nike+ Running. Disponível em: [https://play.google.com/store/apps/details?id=com.nike.plusgps&hl=pt\\_BR](https://play.google.com/store/apps/details?id=com.nike.plusgps&hl=pt_BR). Acesso em: 28 ago 2015.

Noll, M. G., (2011). Running Hadoop On Ubuntu Linux (Single-Node Cluster). Disponível em: <http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-single-node-cluster/>. Acesso em: 14 out 2015.

Olsen, E., Kraft, P., (2009), "ePsychology: A pilot study on how to enhance social support and adherence in digital interventions by characteristics from social networking sites", *Persuasive 2009*

Paschoal Jr, F., Ebecken, N. F. F., (2014), "Uma abordagem para identificação de padrões comportamentais a partir de aplicativos para redes sociais", *Proceedings of the XXXV Iberian Latin-American Congress on Computational Methods in Engineering - CILAMCE 2014*

Paschoal Jr, F., Ribeiro, G. V. S., Daquer, L. M. de A., Mauro, R. C., Ogasawara, E. S., Ebecken, N. F. F., (2015), "FitRank – Desenvolvimento de aplicativo para a identificação de padrões comportamentais de atividades físicas a partir de aplicativos para redes sociais", *Proceedings of the XXXVI Iberian Latin-American Congress on Computational Methods in Engineering - CILAMCE 2015*

RunKeeper, (2015). RunKeeper - The Personal Trainer For Your Pocket. Disponível em: <https://www.facebook.com/games/runkeeper-og/?fbs=133>. Acesso em: 28 ago 2015.

Runtastic, (2015a). Runtastic - makes sports funtastic. Disponível em: <https://www.facebook.com/games/runtastic/?fbs=133>. Acesso em: 28 ago 2015.

Runtastic, (2015b). Runtastic Corrida e Caminhada. Disponível em: [https://play.google.com/store/apps/details?id=com.runtastic.android&hl=pt\\_BR](https://play.google.com/store/apps/details?id=com.runtastic.android&hl=pt_BR). Acesso em: 28 ago 2015.

Runtastic, (2015c). Runtastic PRO GPS Correr. Disponível em: [https://play.google.com/store/apps/details?id=com.runtastic.android.pro2&hl=pt\\_BR](https://play.google.com/store/apps/details?id=com.runtastic.android.pro2&hl=pt_BR). Acesso em: 28 ago 2015.

Runtastic, (2015d). Runtastic Road Bike Ciclismo. Disponível em: [https://play.google.com/store/apps/details?id=com.runtastic.android.roadbike.lite&hl=pt\\_BR](https://play.google.com/store/apps/details?id=com.runtastic.android.roadbike.lite&hl=pt_BR). Acesso em: 28 ago 2015.

Runtastic, (2015e). Runtastic Road Bike PRO GPS. Disponível em: [https://play.google.com/store/apps/details?id=com.runtastic.android.roadbike.pro&hl=pt\\_BR](https://play.google.com/store/apps/details?id=com.runtastic.android.roadbike.pro&hl=pt_BR). Acesso em: 28 ago 2015.

Runtastic, (2015f). Runtastic Mountain Bike GPS. Disponível em: [https://play.google.com/store/apps/details?id=com.runtastic.android.mountainbike.lite&hl=pt\\_BR](https://play.google.com/store/apps/details?id=com.runtastic.android.mountainbike.lite&hl=pt_BR). Acesso em: 28 ago 2015.

Runtastic, (2015g). Runtastic Mountain Bike PRO. Disponível em: [https://play.google.com/store/apps/details?id=com.runtastic.android.mountainbike.pro&hl=pt\\_BR](https://play.google.com/store/apps/details?id=com.runtastic.android.mountainbike.pro&hl=pt_BR). Acesso em: 28 ago 2015.

Silva, A. A. S. da, (2015), *NET-Y: Uma abordagem para detecção de atividades suspeitas em redes sociais*, Universidade Federal do Rio de Janeiro - COPPE Disponível em: [http://www.coc.ufrj.br/index.php/component/docman/doc\\_download/2591-antonio-anddreserpa-da-silva-doutorado?Itemid=](http://www.coc.ufrj.br/index.php/component/docman/doc_download/2591-antonio-anddreserpa-da-silva-doutorado?Itemid=).

The Apache Software Foundation, (2014a). Apache Mahout: Scalable machine learning and data mining. Disponível em: <http://mahout.apache.org/>. Acesso em: 14 out 2015.

The Apache Software Foundation, (2014b). Naive Bayes. Disponível em: <http://mahout.apache.org/users/classification/bayesian.html>. Acesso em: 15 out 2015.

The Apache Software Foundation, (2014c). k-Means clustering - basics. Disponível em: <http://mahout.apache.org/users/clustering/k-means-clustering.html>. Acesso em: 15 out 2015.

The Apache Software Foundation, (2015a). Welcome to Apache<sup>TM</sup> Hadoop®! Disponível em: <http://hadoop.apache.org/>. Acesso em: 14 out 2015.

The Apache Software Foundation, (2015b). Apache HBase – Apache HBase<sup>TM</sup> Home. Disponível em: <http://hbase.apache.org/>. Acesso em: 14 out 2015.

Wikipedia, (2014), *List of social networking websites*. Disponível em: [http://en.wikipedia.org/wiki/List\\_of\\_social\\_networking\\_websites](http://en.wikipedia.org/wiki/List_of_social_networking_websites).

World Health Organization, (2002). World Health Day | 2002: move for health. WHO. Disponível em: <http://www.who.int/world-health-day/previous/2002/en/>. Acesso em: 15 out 2015.

World Health Organization, (2010), *Global recommendations on physical activity for health*.

World Health Organization, (2015). WHO | Physical activity. WHO. Disponível em: <http://www.who.int/mediacentre/factsheets/fs385/en/>. Acesso em: 15 out 2015.