



## **PUBLICAÇÃO CIENTÍFICA DE DOCENTES DA UNIVERSIDADE FEDERAL DE RONDÔNIA**

**ANDRÉ LUIZ DE SOUZA FREITAS**

Universidade Federal de Rondônia

[andre.freitas@unir.br](mailto:andre.freitas@unir.br)

**MARILUCE PAES DE SOUZA**

Universidade Federal de Rondônia

[mariluce@unir.br](mailto:mariluce@unir.br)

### **Resumo**

As pesquisas científicas são testadas e constantemente revisadas para verificação e validação dos dados encontrados. Para tanto, a metodologia empregada e resultados obtidos utiliza variados meios, modos e modelos para verificação da qualidade das pesquisas realizadas nas Instituições Federais de Ensino Superior (IFES). Este artigo tem como objetivo de prospectar tendências e propor padrões que fomentam ou reduzem a capacidade produtiva dos pesquisadores e grupos de pesquisadores. Partindo dos dados coletados sobre as publicações científicas foi possível analisar além dos dados pessoais dos pesquisadores, mas também suas produções científicas, divididas por ano de publicação, tipo de produção e qualificações. Para elucidar os pontos levantados, com o uso da mineração de dados foi montado cenários, a partir de informações gerais sem observar o individual. Porém, durante a análise, foi observado que a ação individual, influencia no geral, seja em um nicho de atuação, seja no total das produções, quando a produção individual é muito grande, podendo haver dois cenários. As publicações com qualidade irão influenciar positivamente no cenário, caso contrário irá depreciar a produção geral, pois quantitativamente o individual influencia na posição geral.

Palavras-Chave: Publicação científica; Mineração de Dados; Universidade Federal de Rondônia.

## 1- INTRODUÇÃO

A comunidade acadêmica, principalmente a área de pesquisa, necessita se manter ativa pesquisando e divulgando seus estudos. Quando a escolaridade se tornou uma função econômica, que busca a melhoria dos produtos, onde a produção de mais produtos em menos tempo, leva a um maior ganho financeiro, a educação, principalmente a superior, toma um caráter econômico/financeiro importante (SAVIANI, 2002). Essa visão mercantilista da ciência trouxe uma nova lógica, onde a qualidade é secundária frente a quantidade focando no produtivíssimo, primando pela quantidade dos trabalhos produzidos e publicados (RODRIGUES, MARTINS, 2005). Nesse contexto, o incremento da produção de artigos científicos leva em consideração o número dos artefatos produzidos.

Mesmo havendo o modelo de publicação ignorando a qualidade da produção, esta deve passar de uma característica implícita, para uma necessidade explícita, não apenas buscando o aumento da quantidade da produção científica, mas também pela qualidade, mudando o paradigma brasileiro, de “*publicar ou perecer*” (SPINAK, 1996).

Mensurar a qualidade de um estudo, o conteúdo de uma pesquisa, não é uma tarefa trivial. A busca por divulgar informações, contribuir com a sociedade, incrementar a cultura e desenvolver uma região através da ciência é almejada pelos pesquisadores e faz parte das diretrizes dos programas de pós-graduação *Stricto Sensu*. Porém, nem a quantidade nem a qualidade da produção podem ser deixadas de lado, uma vez que a produção da ciência é parte do trabalho realizado pelas Instituições Federais de Ensino Superior (IFES).

A necessidade de publicação é um paradigma que conta não somente com a publicação em quantidade, mas também com qualidade (ZAGO, 2011). A preocupação com a qualidade dos estudos publicados não vem sendo observada pelos pesquisadores brasileiros, mas esta está vinculada diretamente a necessidade de aferir a qualidade dos estudos, contudo, a qualidade de um estudo tem inúmeras maneiras de ser mensurada, desde a relevância do estudo apresentado, passando pela sua abrangência, e pela sua utilidade dentro da comunidade acadêmica. Nesse sentido, são criados indicadores nacionais e internacionais.

No Brasil, a base oficial de verificação para a estratificação dos periódicos, visando a classificação da produção intelectual no seio dos programas de pós-graduação brasileiros, é o indexador da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), vinculada ao Ministério da Educação (MEC). Essa estratificação é realizada, de forma indireta, utilizando um índice denominado de Qualis, que estabelece, uma escala de relevância decrescente (A1, A2, B1, B2, B3, B4, B5 e C), para a qualidade dessa produção. Existe também o Qualis livros, porém essa classificação, como diz o nome, gera uma escala para os livros científicos, mas essa escala não será considerada neste trabalho.

## 2 - REVISÃO TEÓRICA

### 2.1. A Publicação Científica

A publicação de artigos em periódicos é o meio através do qual os pesquisadores, expõem suas ideias, garantindo a propriedade científica e se submetem à avaliação de outros pesquisadores, sendo parte essencial do processo de produção do conhecimento científico e da inovação (COIMBRA JR., 1999).

O FI determina a frequência com que um artigo é citado por outros pesquisadores, como uma forma de classificação e avaliação dos artigos. O FI considera a quantidade de

publicação de um determinado periódico, nos últimos dois anos, ou seja, o FI de 2015 é calculado com a somatória dos artigos dos anos de 2013 e 2014 e a quantidade de vezes que um artigo foi citado, então é realizado uma divisão da quantidade de vezes pelo número de artigos publicados, dessa forma, quanto mais citado, mais rígido será o critério para aceitação de artigos em periódicos com maior FI (MARZIALE, *et al.*, 2002).

A publicação de um artigo torna público o esforço para produção de um estudo que, antes de ser publicado e mesmo que não seja aprovado, traz ao autor do estudo informações valiosas para o artefato científico, uma vez que no processo de revisão do artigo está presente a preocupação com a qualidade do periódico, e que o processo editorial distingue a proposta do estudo e a sua contribuição, pois é no processo de revisão que a comunidade ratifica as contribuições e se o mesmo deve ser compartilhado com o restante dos pesquisadores (DAVIS, 2014).

As informações sobre as publicações científicas brasileira é indexada em uma plataforma para agrupar a produção dos pesquisadores que atuam nas instituições de pesquisa no Brasil, essa plataforma é chamada de *Lattes* e está vinculada ao CNPq. A CAPES criou, em 1998, o sistema Qualis, com a finalidade de classificar os periódicos que veiculam a produção científica dentro e fora do país. O Qualis foi concebido durante o programa de informatização da CAPES. Seu objetivo é a classificação para divulgação de periódicos; anais; jornais ou revistas – citados pelos programas de pós-graduação.

Mesmo tendo os indexadores e FI, nem todos periódicos com FI alto estão presentes na base CAPES. Existem, no ano de 2016, mais de 100 mil periódicos indexados, classificados através dos estratos A1, A2, B1, B2, B3, B4, B5 e C. Os quatro primeiros estratos do Qualis estão classificados, considerando o Fator de Impacto, conforme o quadro 1.

Quadro 1 - Equivalência Qualis CAPES X Fator de Impacto

<b>Qualis</b>	<b>FI</b>
<b>A1</b>	Igual ou Maior que 3,8
<b>A2</b>	Entre 2,5 – 3,79
<b>B1</b>	Entre 1,3 – 2,49
<b>B2</b>	Entre 0,001 – 1,29

Fonte: Elaborado a partir de Bastos (2010).

Ainda assim, os critérios de classificação adotados pela CAPES são distintos para as diversas áreas de conhecimento. Para cada conjunto de grupo de cursos são verificadas as publicações registradas no período anterior e realizada uma nova análise pelo comitê de análise de cada área, composto por membros da comunidade acadêmica e convidados da CAPES para realizarem a revisão. O comitê se reúne e faz a revisão conforme os critérios estabelecidos para a área. Dessa forma um periódico A1 da área de Administração, Ciências contábeis e Turismo, tem critérios distintos de um periódico A1 de Medicina I.

## 2.2. Redes de Cooperação para Produção Científica

A busca pela cooperação visa criar um ambiente integrador, acima de conceitos ideológicos, políticos ou de outra natureza (AUMANN, 1964 E SCHELING, 1958). A criação de redes, tanto para cooperação entre indivíduos quanto para a forma como esses atores se relacionam, traz, em sua maioria, benefícios que refletem em toda organização.

As pessoas, em suas relações humanas no âmbito laboral, desejam que cada um desempenhe suas funções da melhor forma, para o seu trabalho e o melhor para o grupo (SILVA, 2008). Na contrapartida, existem indivíduos que não colaboram com outros atores do seu meio, podendo até produzir trabalhos, mas a sua dinâmica de trabalho solitária faz com que seu conhecimento não seja compartilhado, nem seja influenciado ou influencie outros, como por exemplo o ambiente laboral dos pesquisadores (SILVA, 2008).

### **2.3. Teoria da Agência**

A abordagem da Teoria da Agência corrobora para tomada de decisão para mais de um indivíduo, que tem enfoque prescritivo, utilizando as regras notadas dentre os modelos normativos da teoria e das entidades vinculadas ao principal, esperando que o agente cumpra suas atividades remuneradas (HENDRIKEN e BREDÁ, 1999). Neste sentido, as IFES podem ser analisadas a partir dessa perspectiva, uma vez que tem suas particularidades, tendo, em dado momento, a gestão das IFES como o principal e em outro como agente. Mesmo havendo o papel da teoria clássica da empresa, existe a figura do proprietário-administrador, tendo como dicotomia básica a atração no mesmo papel, como o ator que fornece o recurso financeiro e o administrador do mesmo recurso (FERREIRA, VITORINO FILHO, 2010). Embasada nessa dicotomia da gestão clássica, há a sedimentação da teoria da agência, que mostra os problemas clássicos da gestão moderna, que são: conflito da agência e assimetria informacional (LOPES, 2012).

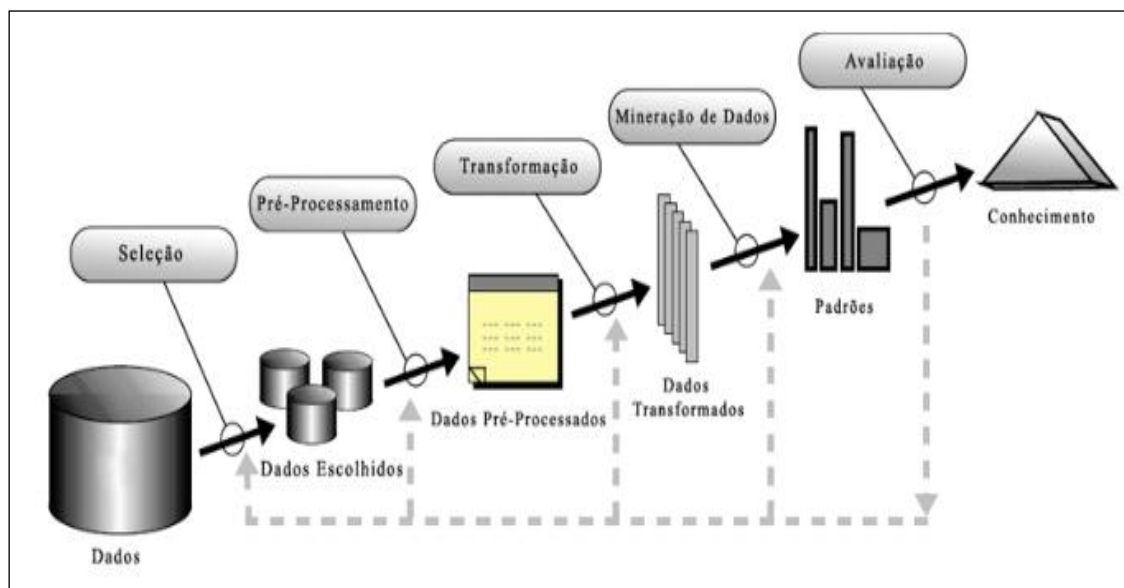
Dentre um dos modelos de controle empregados, a mineração de dados permite a predição de situações que podem limitar a produção científica, assim como a possibilidade de haver um controle em áreas com menor eficácia, tornando assim a possibilidade de haver um controle antecipado das ações dos agentes.

### **2.4. Mineração de dados na identificação de padrões**

Mineração de dados, ou *Data Mining*, é um ramo da computação que teve início na década de 1980, quando as organizações começaram a perceber grandes volumes de dados informáticos estocados e inutilizados internamente (HAN, KAMBER, 2006). Nesse contexto, a Mineração de dados servia basicamente para retirar informações de gigantescas bases de dados de maneira automatizada. Atualmente, o *Data Mining* analisa os dados após a extração, de forma a identificar padrões reais e hipotéticos, extraíndo ou minerando conhecimento de grandes volumes de dados (AMO, 2015). Assim, a mineração de dados é uma das alternativas mais eficazes na extração de conhecimento, pois, a partir de grandes volumes de dados, possibilita a descoberta de relações ocultas, padrões e regras capazes de prever e correlacionar dados com intuito de alcançar um maior grau de confiança do conteúdo extraído (GALVÃO, MARIN, 2009).

O termo "mineração de dados" deriva do termo inglês *Knowledge Discovery in Database* (KDD), que significa a busca por novas informações através da análise de grandes volumes de dados, que visa identificar relações entre os dados que podem produzir novos conhecimentos e gerar novas descobertas científicas que podem favorecer a tomada de decisão (BAKER *et al*, 2011).

Figura 1 – Descoberta do conhecimento (KDD)



Fonte: Fayyad et al 1996.

As tarefas podem ser agrupadas em preditivas e descritivas. As preditivas podem ser de classificação e regressão, a primeira consiste na predição de uma variável categorizada para descobrir uma função que mapeie um conjunto de variáveis já definidas, chamadas de classes, a segunda refere-se a busca de funções lineares ou não, onde a variável a ser predita consiste em um atributo numérico contido em um banco de dados (GALVÃO; MARIN, 2009).

### 3 - METODOLOGIA

Nesse trabalho serão abordados as informações referentes à produção científica no levantamento quantitativo, e mais especializado para os dados dos artigos completos publicados em periódicos, que tem sua análise mais profunda e subjetiva, tendo assim o mote qualitativo do trabalho, sendo assim é uma pesquisa qualitativa, pela sua natureza de análise das redes de relacionamentos entre os docentes e pela sua análise com auxílio da mineração de dados e quantitativa, pelos dados levantados, quanto aos números de publicações, distribuição das publicações por sexo, localidade, nível de instrução e idade.

Deve ser levado em consideração que é essa a dimensão utilizada pelos órgãos de controle do governo federal para analisar a quantidade e qualidade da produção científica dos docentes das IFES, mesmo os artigos completos correspondendo a 16% da produção científica e a 13% da publicação geral. Será utilizada essa variável não somente para a quantidade de artefatos produzidos, como também para classificar a qualidade dos artigos, mensurados pelo indexador da CNPq chamado Qualis. Além da publicação de artigos, existem outras informações presentes no Lattes que foram analisadas, como o gênero dos pesquisadores, localidade, nível de instrução, faixa etária e grau de obsolescência das informações presentes nos currículos dos docentes. Tais informações que serão abordados mais especificamente para os docentes da UNIR, podem também ser aplicados para outras IFES, objetivando a análise da qualidade da produção científica, sendo essa pesquisa de natureza aplicada.

### 3.1. Estudo de Caso

Para elaboração desta pesquisa utilizou-se base quali-quantitativa, composta a partir da análise dos currículos cadastrados na plataforma Lattes do CNPq. A análise dos currículos foi focada em uma série histórica da produção anual, para observar a relação de casualidade entre a existência de pós-graduação, e qualidade dos artigos produzidos. Foram analisadas as produções científicas de 2008 até 2015, por docentes da Fundação Universidade Federal de Rondônia (UNIR). Vale ressaltar que, tal análise não tem a finalidade de qualificar os trabalhos produzidos, uma vez que esta qualificação é realizada pelas organizações nacionais e internacionais, o que é necessário para a criticidade do estudo para favorecer a reflexão sobre a solução dos problemas (BUFREM, 2006).

Desse modo, não foram avaliadas a produção e sim os atributos que contribuem ou não, para uma produção de qualidade dentro da academia segundo os critérios de avaliação já consolidados. Tais atributos estão vinculados aos nichos de atuação dos docentes: NÚCLEOS E CAMPI. Identifica-se a produção, grau de instrução dos pesquisadores e se atuam em programas de pós-graduação *stricto sensu*, bem como, observa-se a casualidade entre o nível de qualificação dos docentes, atuação em algum programa de pós-graduação e a qualidade na produção científica, de modo a analisar todas as publicações de artigos e suas respectivas classificações. O procedimento racional e sistemático proporciona respostas a problemas que são propostos nas pesquisas (GIL, 2002).

O ambiente selecionado para pesquisa é a UNIR, tendo os seus docentes como população. As produções científicas foram analisadas em todas as dimensões, foram utilizadas nas pesquisas e levantamentos quantitativos de produção, o restante da análise foi realizado tempo como finalidade, artigos completos publicados. A população é composta de 848 docentes os quais estão lotados em uma das unidades divididas em núcleos e *campi*. A coleta dos nomes e dos currículos, ocorreu no início de 2016, o intervalo da pesquisa das produções compreende o período de 2008 até 2015.

Registra-se que dentre os currículos considerados, 66 não foram analisados, uma vez que não foram localizados na plataforma Lattes, podendo-se atribuir a questões como: problemas de grafia diferente, abreviação, acréscimo ou supressão de nomes.

O sistema foi inicialmente alimentado por um arquivo de texto que considerou quatro informações diferentes de cada currículo a ser pesquisado: identificação, nome, intervalo de pesquisa e função. No quadro 2 estão descritas a definição operacional considerada.

Quadro 2 – Descrição Operacional das Informações do Sistema

<b>Informação</b>	<b>Descrição Operacional</b>
Identificação (ID)	Código de 16 dígitos que é associado a cada currículo no ato do seu cadastro. É um número único e sequencial livre na plataforma.
Nome	Informar o nome do pesquisador e efetuar uma verificação, pois pode haver uma grafia diferente do informado na lista, primeiro é verificado o Id e depois é verificado o nome.
Intervalo da Pesquisa	O sistema pode ter uma configuração padrão para pesquisar o intervalo entre os anos, como também, ser configurado individualmente.
Função	Informar a função da pessoa a ser pesquisada, que pode ser professor, colaborador, aluno.



Fonte: Dados da pesquisa.

A partir da leitura dos currículos, aqueles que continham as duas primeiras informações: identificação do Lattes e nome, foi realizada a conversão dos dados para deixar em formato de navegado, chamado de html, que permitiu identificar e extrair áreas ou trechos específicos de texto (CHALCO e CESAR-Jr., 2009).

### **3.1.1. Fonte de Dados**

Várias fontes de dados foram utilizadas, sendo predominantemente duas: (1) Sistema Integrado de Gestão Universitária (SINGU) e (2) Plataforma Lattes do CNPq, além de documentos de planejamento estratégico institucional, relatórios gerenciais, portais institucionais e portais dos programas *stricto sensu*. A base de dados do SINGU, serviu de base oficial para formar a lista dos docentes que foram consultados, todos os docentes com vínculo com a instituição no período da pesquisa, atuado ou tendo algum vínculo institucional que o fez ser registrado no sistema acadêmico da UNIR, teve seu currículo consultado. Outra importante contribuição deu-se com a extração dos currículos Lattes dos docentes listados anterior e posterior a análise dos mesmos por uma ferramenta chamada de Script Lattes.

### **3.1.2. Coleta de Dados**

A coleta de dados desse trabalho se deu primeiro pela obtenção da lista dos docentes, posteriormente, acesso a todos os currículos dos docentes listados pelo SINGU. A obtenção dos currículos se deu através da ferramenta disponibilizada pelo portal de busca por nome do CNPq em sua plataforma Lattes, pelo nome de cada docente através de seu sítio eletrônico na internet.

Com os currículos baixados, foi formado um novo repositório local, tendo em vista que, por uma nova regra de segurança da plataforma Lattes, não pode haver consulta sequencial, somente consultas unitárias, o que dificultou bastante a coleta dos currículos. Todos os documentos gerenciais utilizados são de domínio público e estão disponíveis para livre consulta no portal institucional da UNIR ou nos portais institucionais das unidades da própria UNIR.

Com o pré-processamento dos dados utilizando a ferramenta *scritLattes*, partindo do repositório local formado, os dados foram estruturados em planilhas para junção com os dados obtidos nos portais institucionais.

## **3.2. Análise de Dados**

O processo de análise dos dados coletados se deu inicialmente com a identificação dos docentes que estão cadastrados na base de dados do SINGU, utilizando o método de busca estruturada no banco de dados da UNIR. Com base na lista nominal dos professores formulada foi realizada a extração dos currículos da plataforma Lattes.

Os currículos extraídos foram agrupados em um servidor de internet local, para que fosse lido por um sistema chamado *ScriptLattes*, que faz a extração dos dados dos currículos e agrupam de forma estruturada, trazendo os dados sequencialmente de cada pesquisador em linhas de registros, para que pudessem ser analisados de maneira mais simples e rápida.

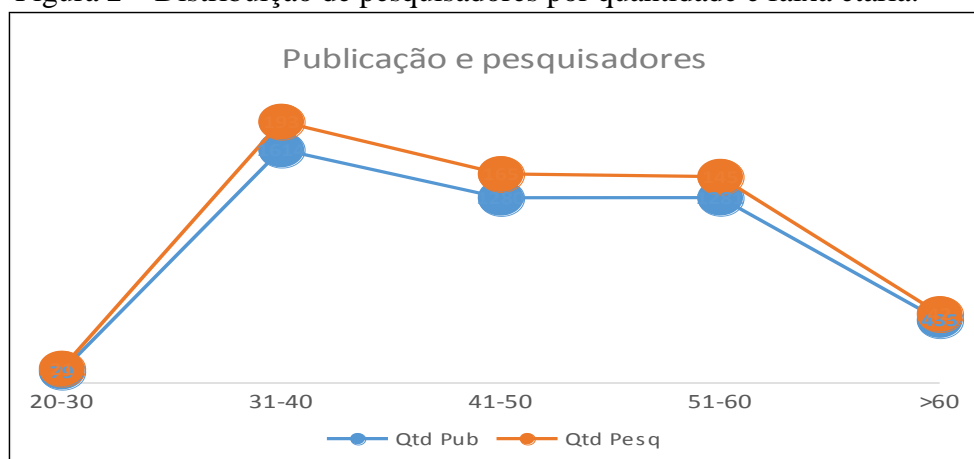
Após lido os arquivos do servidor local de internet, foram gerados os relatórios das análises preliminares da própria ferramenta e arquivos de dados extraídos em formato de texto simples e sem formatação.

## 4 – RESULTADOS E DISCUSSÃO

### 4.1. Identificação da produção científica e técnica dos docentes da UNIR

O levantamento dos dados contidos nos currículos lattes, demonstra uma realidade que pode explicar o grande número de pesquisadores individuais, ou ainda, aqueles que não podem ser contabilizados por falta de registros de sua produção no currículo lattes. Na figura 3, observa-se que 54% dos currículos foram atualizados a menos de 6 meses, embora isto seja uma das metas do Plano de Desenvolvimento Institucional (PDI), bem como, assunto da Instrução Normativa (IN) 01/2014/PPGE/UNIR, que trata da necessidade de atualizar os currículos, assim como as portarias 191/2011 e 192/2012 e a resolução 250/CONSEA. Apesar da IN, portarias, resolução e PDI. Até o início de 2016, quando se fez a coleta de dados para o período de 2008 a 2015 desta pesquisa, quase metade dos currículos permanecem desatualizados. A falta de atualização nos currículos, impedem que a produção docente seja computada, pelos entes da administração pública, impossibilitando que os cursos de graduação e de pós-graduação obtenham melhoras em suas avaliações. Esta situação prejudica o pesquisador e a instituição.

Figura 2 – Distribuição de pesquisadores por quantidade e faixa etária.



Fonte: Dados da pesquisa.

O grande número de publicações na faixa de 31-40, apresentam um quadro promissor a UNIR, tendo em vista que a maioria dos pesquisadores estão nessa faixa, o que favorece à IFES com um quadro de instituição jovem e com produção boa em número e qualidade. No entanto, esta constatação não corrobora com o quadro de obsolescência dos currículos lattes.

A produção computada, no período de 2008 até 2015 nesse trabalho é o que está registrado no Lattes, uma vez que não há outro meio de registrar e divulgar a produção científica que seja considerada na gestão pública. Dentre as informações levantadas, também deve ser observado que do número total de docentes da UNIR, 40% tem registrado produção científica de artigos publicados, os outros 60% não tem registrado suas publicações de artigos.



Figura 3 – Obsolescência do Lattes

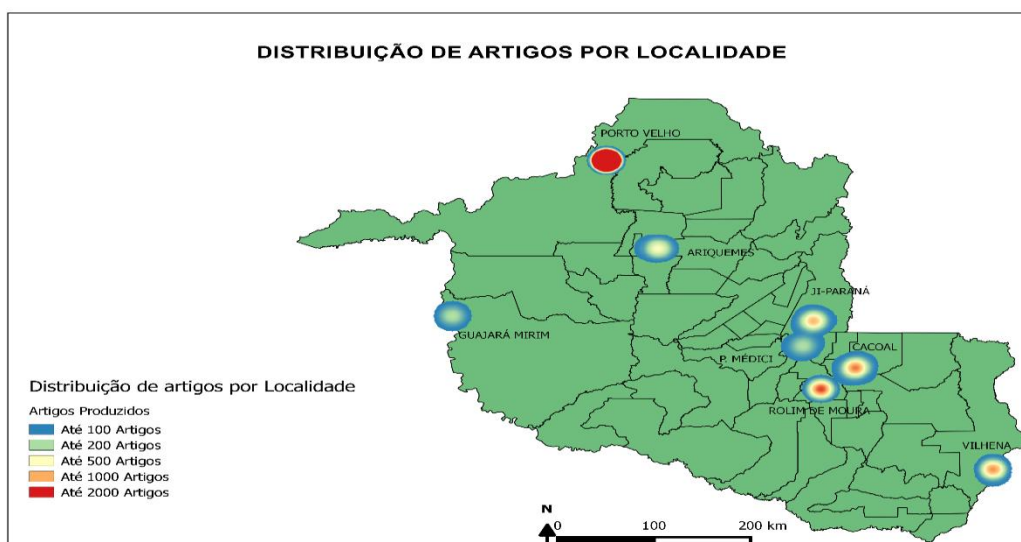


Fonte: Dados da pesquisa.

Observa-se, na figura 2, que 28% dos docentes atualizaram seus currículos entre 6 e 12 meses, que 13% mais que 19 meses e 5% situa-se entre 12 e 18 meses, esta situação torna-se preocupante quando se trata de professores e pesquisadores vinculado a uma IFES. Foram analisados 848 currículos, desses, 564 tem artigos publicados, destaca-se que a distribuição entre as unidades apresenta proporção similar ao percentual de atualização dos currículos.

A distribuição geográfica da produção científica dos docentes dentro do estado, pode ser observada na figura 4, mapa de calor, que demonstra a quantidade de artigos produzidos em cada *campi* da UNIR. Pode-se observar a maior intensidade no campus de Porto Velho e nos *campi* onde, existem programas de pós-graduação *strictu sensu*.

Figura 4 – Mapa de calor de distribuição da produção de artigos dos docentes por Campi.



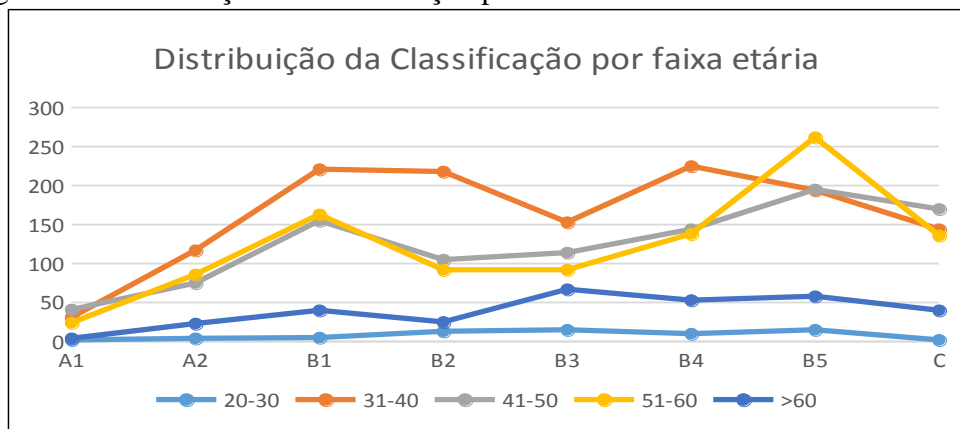
Fonte: Dados da pesquisa

A concentração no campus sede e onde são oferecidos cursos *strictu sensu* é tratado por Waiczky e Ensslin quando se refere a necessidade de divulgação dos trabalhos existentes e produzidos na divulgação da pesquisa, que corrobora com o mapa de calor mais intenso nas localidades que tem presença de cursos de pós-graduação, conforme a figura 4,

apresenta a distribuição das produções por área inclusive as produções técnicas, que devem ser contabilizadas.

A estratificação dos docentes por faixa de atuação e níveis de instrução, fornecem quadro de como são distribuídas as pesquisas pela faixa etária e níveis de instrução, tanto em quantidade de docentes quanto pelo quantitativo de produção por faixa etária, e ainda, a classificação por faixa etária.

Figura 5 – Distribuição da classificação por faixa etária.



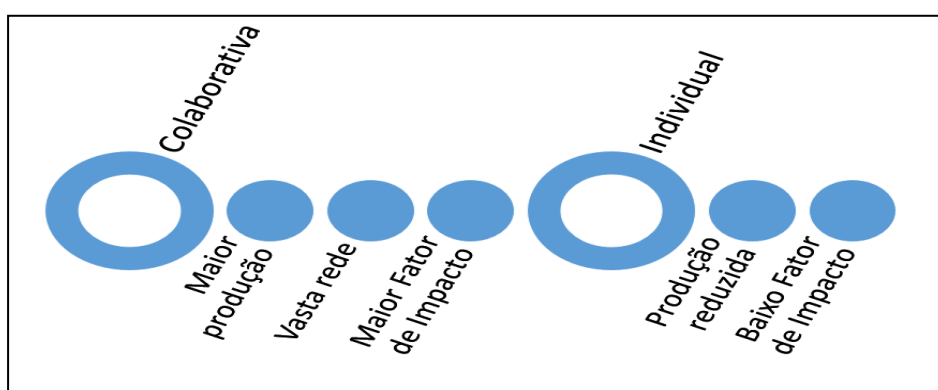
Fonte: Dados da pesquisa.

A quantidade de artigos publicados entre os docentes, segue o quantitativo de pesquisadores por cada faixa, conforme figura 5, os pesquisadores com idade entre 31 - 40 anos, tem maior quantidade de artigos publicados em periódicos com FI. Esses dados dão um quadro diferente da rede de relacionamentos, pois indica que os maiores agrupadores das redes de relacionamento são docentes que estão nas duas últimas faixas etárias, o que poderá ser visto nos itens que seguem.

#### 4.2. Redes de relações a partir da publicação científica

Os dados levantados corroboram com a teoria de redes, as conexões entre os atores, que formam a rede de colaboração, apresentam um quantitativo de produção científica, quando o grau de interação entre os pesquisadores é maior, a rede de colaboração entre os membros, formando nichos de colaboração, criando vários *hubs*, que na teoria de redes, conta com o efeito de centralidade (SILVA, 2008). Alguns dos fatores responsáveis pela rede colaborativa, estão apresentados na figura 6.

Figura 6 – Rede colaborativa e pesquisadores individuais.



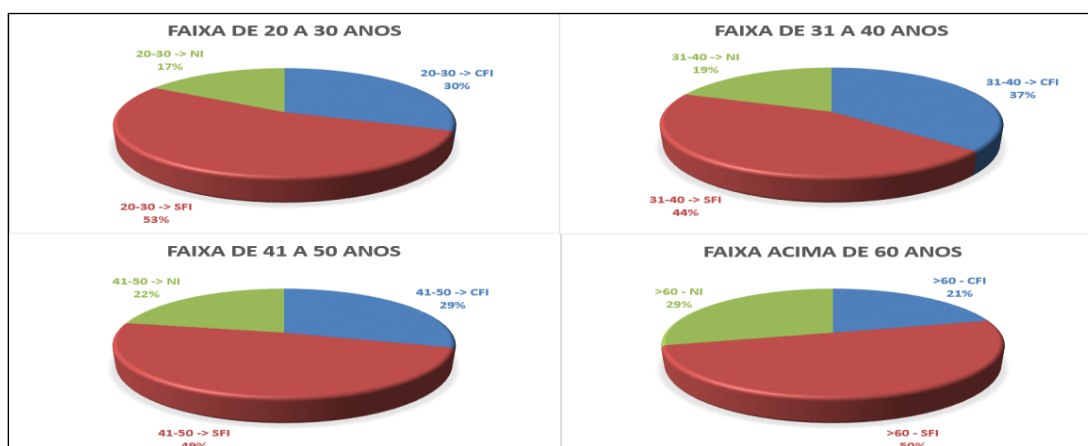
Fonte: Dados da pesquisa.

Quanto maior a rede, surgem inúmeros pontos de centralidades, que tornam a colaboração entre os pesquisadores mais robustas, pois mesmo que haja um ponto de ruptura, as conexões podem tomar outros caminhos e refazer os caminhos de interligação (TRIVERS, 1971).

### 4.3. Tendências da Mineração de Dados

A mineração de dados, tenta trazer dados que antes foram ignorados através de ligações, inferências distintas e cruzamentos de dados que antes foram desprezadas ou não fizeram muito sentido. Os dados obtidos foram formatados para serem submetidos a ferramenta de mineração de dados, o cruzamento dos dados através de alguns algoritmos de descoberta, iniciando com o algoritmo J48, que é um algoritmo derivado do ID3, que busca na base de dados montar uma estrutura, chamada de árvores de decisão por indução, esta árvore é montada, separando parte da base de dados para testar, após a separação da base é verificado se a classificação da base está sendo separada em classes corretas e a separação das categorias de forma homogênea no conjunto de dados, que são características chamadas de entropia e ganho de informação. Ao classificar as informações em classes e categorias, o J48 utiliza o método de divisão e conquista, incrementando assim a capacidade de predição de cenários, existentes dentro da árvore de decisão. Desse modo é possível avaliar o melhor passo a ser seguido localmente, sem um grande cuidado com a produção do melhor cenário, daí o motivo de interpretar a árvore gerada e a partir dela “colher” as melhores regras apresentadas. Uma vez que a árvore gera uma estrutura maior, sub árvores até as folhas, manipulando uma grande quantidade de dados, sendo mais de 4000 registros analisados (AMO, 2015). A mineração é utilizada para descoberta de padrões, contudo os dados foram antes de passarem mineração de dados em busca do KDD, pode trazer informações substanciais já na fase de dados transformados, conforme informações apresentadas na figura 7.

Figura 7 - Produção por faixa etária.



Fonte: Dados da pesquisa.

Seguem as regras encontradas nas estruturas de decisão para cada localidade, vale ressaltar que as quantidades de regras dependem dos cenários, quantidades de instâncias reportadas de maneira correta.

#### 4.4. Um framework para avaliação da produção docente

Mesmo tendo uma grande quantidade de métricas para mensurar a produção científica, tais métricas são aplicadas pelas plataformas indexadoras que concentram os trabalhos científicos produzidos, como a plataforma Lattes e o seu classificador oficial que é o Qualis da CAPES, sendo esse o classificador aceito pelos entes de controle do Governo Federal e entidades de fomento à pesquisa no Brasil.

Tendo esses trabalhos armazenados nessas plataformas, uma vez que, não existe por parte das IFES esses dados estruturados para consulta e referência local, tornando necessário percorrer esse caminho inverso, do registro na plataforma até a IFES à qual o docente está relacionado passando por diversos processos e bases de dados distintas, além de sítios eletrônicos.

A formação dos dados para pesquisa da produção dos docentes, parte da formulação de uma lista dos docentes, esse trabalho focou na produção dos docentes, contudo pode ser mais abrangente, como o corpo administrativo da instituição ou até mesmo os alunos. Essa lista pode ser obtida no sistema acadêmico da IFES, seja um sistema informatizado ou não, mesmo que a IFES não possua um sistema acadêmico informatizado, tem a listagem dos docentes que tem vínculo com a instituição. Dessa lista nominal, é formada uma lista com duas colunas, uma com o nome e outra com o código de 16 caracteres que é o código do Lattes, ou seja, a matrícula de cada pesquisador na plataforma Lattes.

Desse ponto, pode haver três saídas, descritas no quadro 3.

Quadro 3 – possibilidades de consulta dos currículos.

Possibilidade	Descrição
CNPq normal	Portal do CNPq normal, consultando os currículos sem verificação de interação humana
CNPq Capcha	Portal do CNPq solicitando verificação de interação humana, dispositivo chamado de capcha que força o usuário a confirmar uma numeração apresentada na tela.
Base local	Convênio entre a IFES e o CNPq, para disponibilização à IFES a base de dados dos currículos registrados na plataforma Lattes, para ser consultada localmente.

Fonte: Dados da pesquisa.

Uma vez que essa lista com as duas colunas esteja completa, deve ser utilizada uma ferramenta para leitura dos currículos, a leitura deve estar direcionada para o código XML (Linguagem utilizada na Internet para montar páginas de forma estruturada), pois a página da plataforma Lattes é construída utilizando XML, o que facilita a categorização da informação, pois os dados na página, são estruturados em páginas de internet que usam XML, além de permitir o download dos currículos (MENA-CHALCO, CESAR JUNIOR, 2011).

Com a estratégia definida sobre a maneira de consultar os currículos, foi selecionada a ferramenta scriptLattes, para fazer a leitura. O scriptLattes faz uma varredura e agrupamento das informações, gerando um arquivo texto, arquivos texto são arquivos sem formatação, com as informações em texto simples, porém com delimitador de campo, pois elas são separadas por; um campo do outro. Então o arquivo em texto é transformado em arquivo separado por vírgula (CSV), onde já está estruturado. Essa etapa é chamada de pré-processamento, pois os dados estão tabulados, mas não traduzem informações úteis.

Nessa etapa, mesmo com a geração do arquivo texto, também geram muitas outras informações, como quantitativo da produção científica, não só de artigos produzidos, mas também de livros, periódicos, resumos e eventos, assim também como produção técnica. O scriptLattes agrupa várias informações, em sua maioria dados quantitativos. Porém, o scriptLattes, também faz uma referência com todos os pesquisadores, gerando uma estrutura matemática chamada de grafo, gerando interligações entre os autores, inclusive com peso para as relações.

Com os dados estruturados, em formato CSV, são pré-processados novamente com a finalidade de obter um arquivo resultante no formato ARFF, que é o tipo de arquivo que pode ser processado pela ferramenta de mineração chamada de Weka (Waikato Environment for Knowledge Analysis), ferramenta para mineração de dados desenvolvida na Universidade de Waikato na Nova Zelândia. Ela foi licenciada sobre a Licença Pública Geral (GPL, acrônimo em inglês). O uso da Weka foi para processar o arquivo ARFF, como a Weka é uma ferramenta que agrupa diversos algoritmos para predição de padrões e é utilizado na mineração de dados, desse modo os arquivos gerados são tendências de padrões encontrados, que nesse trabalho foi utilizado especificamente os algoritmos J48 para árvore de decisão e outro de associação, que foi o Apriori (GALVÃO; MARIN, 2009). Os passos seguidos para chegar ao resultado estão apresentados na figura 8.

Figura 8 - passos do processo para obtenção da informação.



Fonte: Dados da pesquisa.

## 5 – CONSIDERAÇÕES FINAIS

As informações apresentadas nessa dissertação, obtidas após a análise dos dados de diversas fontes, como plataforma Lattes, SINGU e portal institucional da UNIR, e de todos os programas de cursos stricto sensu da UNIR, mostram que grande parte da produção científica e técnica da UNIR, não é considerada, pois em torno de 13% do total do trabalho laboral dos docentes em pesquisa, são de artigos publicados em periódicos.

A discussão a respeito da produção dos pesquisadores, vai além dessa dissertação, mas certamente é algo a ser considerado pelos entes da administração pública, que devem também considerar, conceituar e classificar as outras formas de produção e divulgação de difusão do conhecimento gerado pelas IFES.

O cenário dos pesquisadores da UNIR, apresentado pelos dados, foi que a UNIR tem produção ascendente de artigos, dentre os períodos pesquisados. Em todos os anos houve incremento na produção, menos no ano de 2012, onde houve um decréscimo na produção de artigos, para elucidar essa informação, foram pesquisadas mais de 1800 notícias, chegando ao ponto das greves ocorridas entre os anos de 2010 à 2011, parece que tais fatos refletiram na produção científica no ano de 2012.

Outra informação levada em consideração é a grande quantidade de artigos publicados em periódicos não indexados na plataforma Qualis da CAPES, tais publicações

sem ser indexado, não contam para a instituição, uma vez que os artigos não recebem classificação, voltamos ao um dos pontos desse trabalho, a publicação em periódicos não indexados, contribuem para o quantitativo produzido e sua qualificação não é verificada, ressalto que o fato de não haver indexação na plataforma Qualis, não implica diretamente que o periódico não tenha fator de impacto, não está indexado na base de dados da CAPES.

Quanto aos relacionamentos os resultados apontam para uma grande rede, interligando diversos docentes da UNIR de diferentes cursos e departamentos, tais redes contam com aglutinadores, que são chamadas de *hubs*, esses *hubs*, são docentes que tem publicações com diversos outros docentes de cursos e núcleos diferentes, o que é bom, pois mais de 80% das publicações são compartilhadas entre docentes. Contudo, os principais hubs da rede já estão nas duas últimas faixas etárias, o que causa certo temor, uma vez que com a aposentadoria desses aglutinadores, pode haver uma redução na produção.

Os dados levantados pelo currículo, mostram a divisão dos docentes e pesquisadores, pois nem todo docente tem publicação, o que não poderia ser uma ação dissociada do papel de docentes de uma IFES, assim como a distribuição dos docentes e pesquisadores em cada núcleo e *campi* da UNIR. Uma informação que deve ser observada é o grande número de docentes que não tem seus currículos na plataforma Lattes atualizados. O que pode impactar no número de publicações cadastradas.

A UNIR é uma instituição jovem, pois a maioria dos docentes estão na faixa de 31 a 40 anos, assim como a maioria das publicações, que são acompanhadas pela qualidade na publicação, pois o número de publicações com FI, estão nessa faixa etária. Os docentes nas faixas seguintes devem ser mais, suas publicações acompanham o número de pesquisadores em cada faixa, devido aos concursos quem vem sendo realizados desde a implantação do Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais – REUNI, que foi instituído pelo Decreto nº 6.096, de 24 de abril de 2007, e é uma das ações que integram o Plano de Desenvolvimento da Educação (PDE). Uma informação curiosa, foi que a experiência dos docentes não incrementou o número de publicações em periódicos com FI, tal inferência está mais relacionada ao nível do pesquisador, uma vez que os pesquisadores doutores, tem mais publicações em periódicos com FI.

Apesar da grande quantidade de informações trazidas pelas análises preliminares, a mineração de dados apresenta outras informações, fazendo relações que não foram observadas anteriormente, desse modo foram analisadas por dois tipos de algoritmo, um para montagem de uma árvore de decisão e outro de associação. Tais informações mostraram cenários não presentes na análise dos dados sem a ajuda da mineração, como a produção de dados pelos docentes com doutorado, seja do sexo masculino ou feminino. Assim como é verdadeiro que os docentes que atuam nos níveis de pós-graduação têm publicações com fator de impacto.

A UNIR tem uma grande rede de relacionamento que propicia a colaboração entre os docentes, porém deve sofrer renovação em breve, uma vez que os grandes aglutinadores dessas redes estão próximos à das suas aposentadorias. No entanto, é algo que merece uma atenção por parte das unidades e gestores da UNIR, a fim de garantir a produção científica crescente, observando os pontos de sucesso e os pontos de gargalo que devem ser tratados para incentivar os grupos menos produtivos a produzirem mais, divulgando o resultado dos seus estudos e atualizando seus registros. Com isso, a UNIR pode, inclusive, atender as metas de criação dos cursos *stricto sensu*, propostos no PDI 2014-2018.

## REFERÊNCIAS

- AMO, S.; DIALLO, M. S. ; DIOP, C. T. ; Giacometti, A. ; SOULET, A. **Contextual preference mining for user profile construction**. Information Systems (Oxford), v. 49, p. 182-199, 2015.
- AUMANN, R. J. e MACHAEL, M. **The Bargaining Set for Cooperative Games**. Princeton University Press, Princeton, 1964.
- BAKER, Ryan Shaun Joazeiro; ISOTANI, Seiji; CARVALHO, Adriana Maria Joazeiro Baker de Carvalho. **Mineração de Dados Educacionais: Oportunidades para o Brasil**. Revista Brasileira de Informática na Educação, v. 19, n. 2, 2011.
- BUFREM, L. S. **Linhas e tendências metodológicas nas dissertações do mestrado em ciências da informação do Instituto de Informação em Ciência e Tecnologia - Universidade Federal do Rio de Janeiro (1972-1995)**. Em: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 9, 27 out. a 01 novo de 1996, Curitiba. Anais... (Parte 4.4). Curitiba: UFPr: PUE. 1996.
- CAPES – **Coordenação de Aperfeiçoamento de Pessoal de Nível Superior**. Qualis Capes. Disponível em: <<https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/veiculoPublicacaoQualis/listaConsultaGeralPeriodicos.jsf>>. Acesso em 15 mai. 2015.
- COIMBRA Jr. CEA. **Produção científica em saúde pública e as bases bibliográficas internacionais**. Cad Saúde Pública 1999; 15:883-8.
- DAVIS, G. Editorial essay: **Why do we still have journals?** Administrative Science Quarterly, v.59, n.2, p.193-201, 2014.
- FAYYAD UM, Shapiro GP, Smyth P, Uthurusamy R. **Advances in knowledge discovery and data mining**. Menlo Park, Calif.: AAAI Press: MIT Press; c1996; 611p.
- FERREIRA, Manoel Portugal. **Periódicos e Rankings de Periódicos em Administração**. Revista Contemporâneo em Administração, v. 9, n. 2, Abr./Jun. 2015.
- FERREIRA, Toniel; VITORINO FILHO, Valdir Antônio. **Teoria de Redes: Uma Abordagem Social**. Revista Conteúdo. Capivari, v.1, n.3, Jan./Jul. 2010.
- GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 4 ed. São Paulo: Atlas, 2002.
- GALVÃO, Noemi Dreyer; MARIN, Heimar de Fátima. **Técnica de mineração de dados: Uma revisão da literatura**. Acta Paul Enferm, 2009.
- LOPES, A. B. (2012). **Teoria dos contratos, governança corporativa e contabilidade**. In A. B.



HAN, J., KAMBER, M., **Data Mining: Concepts and Techniques**, Morgan Kaufman, Segunda Edição, 2006.

HENDRIKSEN, Elton S.; BREDA, Michael F. Van. **Teoria da Contabilidade**. São Paulo: Atlas 1999.

MARZIALE, M. H. P.; RODRIGUES, C. M. **A produção científica sobre os acidentes de trabalho com material perfuro cortante entre trabalhadores de enfermagem**. Revista Latino-Americana de Enfermagem, Ribeirão Preto, v. 12, n. 1, p. 571-577, jan./fev. 2002.

MENA-CHALCO, Jesus Pascual; CESAR JUNIOR, Roberto M. **Scriptlattes: Um Sistema de Software Livre para Extração de Conhecimento da Plataforma Lattes**. Jornal da Sociedade Brasileira de Computação, Rio de Janeiro, v. 15, n. 4, p. 31-39, 1 dez. 2009.

RODRIGUEZ, M. V; MARTINS, L. G. A. **As políticas de privatização e interiorização do ensino superior: Massificação ou democratização da educação brasileira**. Revista de Educação, Valinhos, v. 8, n. 8, p. 41-52, 2005. Disponível em: <[www.pgskroton.com.br/seer/index.php/educ/article/download/2210/2105](http://www.pgskroton.com.br/seer/index.php/educ/article/download/2210/2105)>. Acesso em: 17 out. 2015.

SAVIANI, D. **Escola e Democracia**. Campinas: Autores Associados, 35. ed., 2002.

SILVA, Danielly Magalhães da, NUNES, Leandro de Azevedo, ARAGÃO, Nelma Araújo, JUCHEM, Dionise Magna. **A importância do Relacionamento Interpessoal no Contexto Organizacional**. V Congresso Virtual Brasileiro de Administração - CONVIBRA, 2008.

SPINAK, E. **Dicionário enciclopédico de bibliometria, cienciometria e informetria**. Caracas: UNESCO; CII/II, 1996.

TRIVERS, R. L. **Parental investment and sexual selection**. In: B. Campbell (org.), *Sexual selection and descent of man, 1871-1971*. Chicago, Aldine, p. 136-170.

UNIR, **Plano de Desenvolvimento Institucional 2014 – 2018**, UNIR, 2014.

ZAGO, Marco Antônio. **Perfil da Produção Científica Brasileira**. Palestra proferida Inovações Tecnológicas no Brasil FAPESP. São Paulo - SP, em 29 de junho de 2011.